NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

APPLICATION OF ADAPTIVE CERS TO THE KOREA HELICOPTER PROJECT

by

Jaecheon Oh

December 2009

Thesis Advisor:	Dan Nussbaum
Second Reader:	Sung Jin Kang

Approved for public release; distribution is unlimited

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.			
1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** December 2009	**3. REPORT TYPE AND DATES COVERED** Master's Thesis	
4. TITLE AND SUBTITLE Application of Adaptive CERs to the Korea Helicopter Project			**5. FUNDING NUMBERS**
6. AUTHOR(S) Jaecheon Oh			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. government.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited			**12b. DISTRIBUTION CODE** A

13. ABSTRACT (maximum 200 words)

This thesis develops new models to estimate the cost for a defense acquisition project, namely the Korean Helicopter Program (KHP). The thesis constructs various cost estimating models based on the traditional Ordinary Least Square (OLS) method and the Adaptive Cost Estimating Relationships (CER), which was introduced in June 2008. This new methodology is used to improve the uncertainty of OLS as shown in the differences between actual data and predicted values. In particular, the new (Adaptive) CER method uses three ways of estimation to diminish the errors; a priori, piece-wise, and X-distance methods. Among these three approaches, this thesis deals with the priori method which assigns weights to individual data points. By comparing the OLS and the weighted methods, improvements in the cost estimates can be achieved. In addition, this thesis provided robust cost estimates for the KHP.

14. SUBJECT TERMS : Defense Acquisition, Korea Helicopter Program(KHP), Korea Utility Helicopter(KUH), Adaptive Cost Estimation Relationships(CERs)			**15. NUMBER OF PAGES** 99
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

APPLICATION OF ADAPTIVE CERS TO THE KOREA HELICOPTER PROJECT

Jaecheon Oh
Captain, Army, Republic of Korea
B.Eng., In-Ha University, 2000

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
December 2009

Author: Jaecheon Oh

Approved by: Dr. Daniel Nussbaum
Thesis Advisor

Dr. Kang, Sung Jin
Second Reader

Robert F. Dell, PhD
Chairman, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

This thesis develops new models to estimate the cost for a defense acquisition project, namely the Korean Helicopter Program (KHP). The thesis constructs various cost estimating models based on the traditional Ordinary Least Square (OLS) method and the Adaptive Cost Estimating Relationships (CER), which was introduced in June 2008. This new methodology is used to improve the uncertainty of OLS as shown in the differences between actual data and predicted values. In particular, the new (Adaptive) CER method uses three ways of estimation to diminish the errors; a priori, piece-wise, and X-distance methods. Among these three approaches, this thesis deals with the priori method, which assigns weights to individual data points. By comparing the OLS and the weighted methods, improvements in the cost estimates can be achieved. In addition, this thesis provided robust cost estimates for the KHP.

T\HIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

The development and emergence of advanced technology and weapons systems have driven significant increases in defense budgets. On the other hand, limited defense budgets require that resources be used efficiently and effectively. For these reasons, robust, professional and credible cost estimating and analyses are becoming more important for any defense acquisition program.

The Republic of Korea Army (ROKA) has been developing the Korea Utility Helicopter (KUH) since 2005. While some initial cost estimates were developed, they need to be updated in light of new requirements and schedules.

For this reason, the author developed the new CER for the KUH by using traditional Ordinary Least Square (OLS) and Weighted Least Square (WLS) with the Adaptive CER method. Though the traditional OLS method can be used and applied to the KUH, it is difficult to predict the appropriate cost because there is not enough historical and cumulative experience and data for helicopter development in Korea. The new method, Adaptive CERs, was used for the KUH cost estimation in order to overcome these weaknesses.

Military helicopter data was collected through open sources. The ranges of data are main system level, purpose, dimension, weight, and performance. Eight kinds of helicopters were examined to find more feasible data. Furthermore, eight kinds of cost methods, which consisted of one and two variables, linear and power regression, and OLS and WLS were tested. After that, 90 estimates from OLS and 22 estimates from WLS were analyzed. As a result, 28 cost models which are applicable to the KUH were built.

By examining various conditions and methods, the author found that adaptive CER methodology can provide a more stable prediction of cost for the KUH than OLS or WLS alone.

The author presents this new method as a trial for Korea to construct and accumulate the CERs. This new method of cost estimation can be applied to the KUH, as

well as to the Korea Attack Helicopter (KAH) with the use of the cumulative data and experience of the KUH. Furthermore, this method is expected to be used in other defense acquisition projects. The trial, described in the thesis, should contribute to the efficient and effective usage of Korea's defense budget by providing the means for accurate cost estimation.

ACKNOWLEDGMENTS

First of all, I would like to thank my advisors, Dr. Daniel Nussbaum of the Naval Postgraduate School and second reader, Dr. Kang, Sung Jin of the Korea National Defense University (KNDU), for their efforts to lead me in the right way with their powerful knowledge, wisdom and open minds. It has been my great privilege to have such a great advisor and second reader.

Next, I want to express my gratitude to all of the Operations Research faculty members who guided me to develop understanding of OR methodology. If I had not received their devoted teaching and seen their passion for the academy, it would be have been difficult to finish this demanding course.

I want to express my appreciation to the Republic of Korea Army and the Korean National Defense University for providing me the opportunity to achieve a graduate degree at the prestigious Naval Postgraduate School.

I would like to send my thanks to the faculty members and students in Operations Research at KNDU. They taught me the foundations of OR and gave me inspiration to concentrate on study. I want to express my acknowledgment of Professor Lee, Jaeyoung, and classmates of the C4I laboratory, for encouraging me. I appreciate Dr. Park, In-kyoung, for interesting me in cost analysis and teaching me how to study with discipline.

I cannot forget the help of the cost estimating laboratory at KNDU for providing me the data and study materials for this thesis; Major Lee, Yongbok, Ph.D. candidate and Cpt. Eo, Wonjae, Master's Degree candidate at KNDU..

Next, my appreciation is to my classmates at NPS for helping me become accustomed to living and studying in the USA. LT. Gift, Philip (U.S. Navy) and his family are good friends, and he was my academic partner. To Maj. Jung, Bokman, Cpt. Baek, Donghwan, Lee, Wooyoul, Lee, Hyun and Kim, Kangmin, Lt. Kim Sangmin, Koo, Boncheul and Ahn, Hyojoo and other Korean officers who studied with me, all of you are unforgettable people.

Finally, I would like to express my best gratitude to my family for supporting and cheering me on during my NPS life.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. BACKGROUND AND PURPOSE OF THE STUDY

The Republic of Korea Army (ROKA) initiated the Korea Multi-role Helicopter (KMH) acquisition program in September 2001 to provide substitutes for existing helicopters: 500MDs, UH-1Hs and AH-1Ss. The advanced types of helicopters will be usable in combat, light attack, command and control, liaison and passenger-carrying roles. By 2003, the KMH program was under the control and execution of Korea's Agency for Defense Development (ADD) and Korean Aerospace Industries (KAI). However, in 2004, the Korean government required a re-evaluation of the cost of the project, as actual costs became known.

As a result of this reevaluation, the KMH project was cancelled due to the conflict between cost estimating and budget constraints in 2004. However, it was replaced by the less ambitious Korean Helicopter Program (KHP) to develop, at first, a purely utility version helicopter, and later, an attack version based on the utility version. The attack version will be developed, after obtaining additional funding, around 2008–2012.[1]

This helicopter program is very important for the twenty-first century ROK execution of military and civil operations in the Korean environment.

Two hundred and forty-five of this new utility version known as the Korean Utility Helicopter (KUH) are expected to be produced. This program started in June 2006 and has been divided into six phases, as follows: (1) project definition (2006); (2) program development and production of four prototypes (2007–08); (3) prototype ground

[1] KAI SURION, Wikipedia, http://en.wikipedia.org/wiki/KAI_Surion (Accessed July 28, 2009); 한국형 헬기사업(Hankookhyung Helgisaup), Wikipedia,http://ko.wikipedia.org/wiki/%ED%95%9C%EA%B5%AD%ED%98%95_%ED%97%AC%EA%B8%B0_%EC%82%AC%EC%97%85 (Accessed July 28, 2009).

tests (2009); (4) prototype flight tests (2009-11); (5) certification, military standardization and initial production (2010-11); and (6) series production launch (2012).[2] At each phase of the Korean acquisition process, a cost estimate has been required.

In July 2009, the first prototype KUH named SURI-ON was produced. Test flights and operational tests began at that time.

Cost Estimating Relationships (CERs) are the preferred mechanism for predicting the cost of future programs. They are based on historical data of technical and performance characteristics of analogous programs. Regression analyses are the preferred mathematical tool for developing CERs. However, in the case of the KMH, there were not enough data and historical experience with analogous programs to permit development of CERs by those responsible for program management namely the Agency for Defense Development (ADD) and the Defense Acquisition Program Administration (DAPA).

The conflict between cost and budget had an effect on national security and policy. First, the duration of the program was extended at least one more year and a longer time for the helicopter to be deployed into the force will be required. Second, the capability of the weapons system has been downsized in comparison to the requirements in the Requirements of Customer (ROC).

Therefore, it is important to predict appropriate estimated costs

- to prevent the waste of budgeted resources;
- for better alignment of national policies and program execution;
- for better development and justification of the budget; and
- for enhanced stewardship of financial resources.

[2] KAI Surion, *Jane's All the World's Aircraft*, http://search.janes.com/Search/documentView.do?docId=/content1/janesdata/yb/jawa/jawaa333.htm@current&pageSelected=allJanes&keyword=kuh&backPath=http://search.janes.com/Search&Prod_Name=JAWA& (Accessed November 25, 2009).

Credible forecasting of costs is needed to carry out the ROKA programs. This type of forecasting will increase the efficiency of the limited budget and diminish the risk of budget overruns.

This thesis will develop new CERs for KHP based on applying Adaptive-CERs originally developed by Stephen A. Book, Melvin A. Broder, The Aerospace Corporation, and Daniel I. Feldman.

B. ADAPTIVE CER METHODOLOGY

Traditional development of cost-estimating relationships (CERs) has been based on “full” data sets consisting of all available cost and technical data associated with a particular class of products of interest, for example, components, subsystems or entire systems of satellites, and ground systems.[3]

The Adaptive CER is an extension of the concept of “analogy estimating” to “parametric estimating” CERs that are based on specific knowledge of individual data points that may be more relevant to a particular estimating problem than would the full data set. The goal of adaptive CER development is to be able to develop and apply CERs that have smaller estimating errors and narrower prediction bounds. Book’s paper in Appendix A provides a full description of Adaptive CER Methodology.

The Adaptive-CER approach incorporates the following three methods:

First, the A Priori method, which weights each data point by quality or confidence, prior to producing a new CER.

Second, the Piecewise CER method, which groups data into separate subsets which produces small sets of CERs which are more responsive to the value of the independent variable.

[3] Stephen A. Book, Melvin A. Broder and Daniel I. Feldman, “Statistical Foundations of Adaptive Cost-Estimating Relationships,” SCEA(Society of Cost Estimating and Analysis)-ISPA(International Society of Parametric Analysts) Joint Annual Conference & Training Workshop, June 24–27, 2008, 1.

Third, the “X-Distance” method, which weights data points by distance from a cost-driver value of interest and which, therefore, provides analogy-like estimating near the x value chosen.[4]

This thesis will implement only the A Priori method in developing CERs to estimate the cost of the KUH program.

[4] Stephen A. Book, Melvin A. Border and Daniel I. Feldman, “Adaptive Cost-Estimating Relationships”, SCEA(Society of Cost Estimating and Analysis)-ISPA(International Society of Parametric Analysts) Joint Annual Conference & Training Workshop, June 24-27, 2008), 2–3.

II. BACKGROUND

A. PROBLEM STATEMENT

The emergence of new technologies and weapons systems have caused ROKA's defense budget to undergo dramatic increases. However, the defense budget is constrained and it must be utilized efficiently and effectively. Therefore, as the ROKA develops and acquires more KUH, appropriate professional cost estimates will be needed.

Cost estimation and analysis is very important for government acquisition programs for many reasons including: to support funding decisions, to evaluate resource requirements at key decision points, and to develop performance measurement baselines.

Parametric cost models have been utilized worldwide as a means to develop cost estimates as part of larger decision-making processes. However, previous cost models, which were developed in the United States, have limitations when applied to cost estimates in the Korean defense environment.

It is important for Korea to develop its own CERs based on data from its historical experiences in developing and building helicopters. These CERs, when developed, will be used to generate professional, credible cost estimates for current and future acquisition projects. In support of this objective, there has been some research on Korean CER development, not only for helicopters, but for other weapons systems as well. Currently, Korean cost models are being developed, and this thesis is part of that effort.

Nevertheless, little prior data is available, either because it is classified or proprietary. Therefore, this thesis collects and uses only open-source data related to already developed, and similar purpose, helicopters.

B. REVIEW OF PREVIOUS STUDIES

There are two previous studies on the general topic of Korean helicopters.

1. Korean Multi-Purpose Helicopter

Initially the PRICE Suite of Models was used to estimate development and acquisition costs of the KMH. From these results, it was decided to focus first on a Korea Utility helicopter (KUH) and later on the Korean Attack Helicopter (K.A.H).[5] This study is available only in Korean, and it is not included in this thesis.

2. Korea Utility Helicopter (K.U.H) Cost Estimation Report

This report provided initial cost estimates on the KUH to the Korea Defense Acquisition Program Administration (K-DAPA). This study also is available only in Korean, and it is not included in this thesis.

C. ORDINARY LEAST SQUARES (OLS) METHOD

Ordinary least squares (OLS) method minimizes the sum of squared errors between the original dependent variable, y, and the estimated value, $\hat{y}$. If, for example, $\hat{y}$ is modeled by a simple linear equation, namely $\hat{y} = a + bx$, then OLS solves the optimization problem:

$$e_k = y_k - (a + bx_k) = y_k - \hat{y}_k = residuals$$

$$\sum (y_k - \hat{y}_k)^2 = minimum$$

$$\hat{y}_k = \boldsymbol{a + bx_k}$$

The OLS regression method is used to find "best" fits to a set of data points (x_k, y_k)

$$y_k = a + bx_k + e_k, \text{ where } e_k \text{ is } N(0, \sigma^2)$$

[5] Sungjin Kang, Gyumyung Choi, Jongbok Jung, and Seungsoo Kim, "KMH Cost Analysis Report," Korea National Defense University (KNDU) Report for Korea DAPA, December 2005.

- xk is the cost driver and yk is the actual cost;
- ek is the random error between actual cost and estimate;
- $\hat{y}$ k is the predicted cost.

D. ADAPTIVE CER METHOD

The parametric cost-estimating method, also called a Cost Estimation Relationship (CER), can be used to predict the future cost of projects in any phase of its life cycle. CERs are based on historical data and developed using OLS.

Some existing CER methods are influenced by outliers, which can affect the resulting estimates. There are potential ways to address these problems, such as power regression or by using a quadratic method.

The objective of an adaptive CER is to make CERs with more accurate estimating methods, which diminish the estimating errors. The adaptive CER method uses three approaches:

(1) The A Priori method: weighting each point by its quality or the confidence in its accuracy

(2) The Piecewise CER method: grouping data into separate subsets based on natural values of interest

(3) The X-distance method: Weighting points by distance from a cost-driver value of interest. [6]

This thesis will implement only the A Priri method in developing CERs to estimate the cost of the KUH program.

(1) A Priori Method

Book, Broder and Feldman (2008) described the A Priori method this way:

> This method focuses on statistical foundations of the derivation of adaptive CERs, namely the method of weighted least-squares (WLS) regression. Ordinary least-squares (OLS) regression has been traditionally applied to historical-cost data in order to derive additive-error CERs valid over an entire data range, subject to the requirement that all data points are weighted equally and have residuals that are distributed according to a common normal distribution. The idea behind adaptive CERs, however, is

[6] Book, Border and Feldman, "Adaptive Cost-Estimating Relationships," 2–3.

that data points should be "de-weighted" based on some function of their distance from the point at which an estimate is to be made, i.e., each historical data point should be assigned a "weight" that reflects its importance to the particular estimation that is to be made using the derived CER.[7]

[7] Book, Broder, and Feldman, "Statistical Foundations of Adaptive Cost-Estimating Relationships," 5–6.

III. DEVELOPING THE KUH CERS WITH AN ADAPTIVE CER

A. DEVELOPING THE METHODOLOGY

In conducting this research, the author collected, normalized and analyzed helicopter data, and found some significant cost drivers at the helicopter system level. These steps are described more fully in the paragraphs below.

1. Data Collection

All data were collected through books and open sources, such as *JANE's All The World's Aircraft*. Some data was obtained from the Korea National Defense University (KNDU).

2. Data Normalization

All cost data were normalized to $FY08, using NCCA Inflation Indices, available at http://www.ncca.navy.mil/services/inflation.cfm. All technical data were converted to metric specifications.

3. Data Analysis

The author compared OLS-based and WLS-based equations to estimate the cost relationship and developed Adaptive CERs using the WLS method. This research is thought to be a first attempt of its kind, and it is meaningful in terms of developing a CER to estimate the average unit production cost for the KUH, using historical costs and physical characteristics in a Korean development environment.

B. DATA COLLECTION

1. Data Collection

Historical data on helicopter development is difficult to obtain, either because of security or proprietary concerns. Instead, the author collected data from open sources.

The main source of data was Jane's *All The World's Aircraft*. Other data sources are listed in the Reference section. The only Korean helicopter development data available was in the 2004 KMH cost analysis.

Table 1 displays the data collected for this thesis. There are eight helicopters, each with nine descriptive variables. KUH data are not going to be included to the regressions.

Table 1. Collected Helicopter Data

Name	Type	Unit cost (FY08$M)	Weight(kg)			Power Plant	Dimensions (m)		Speed (km/h)		Range (km)
			Empty	Max Taking-Off	Max disc loading	SHP	Main Rotor	Height	Max	Cruise	Max
KUH	Utility	14.10	4,923	8,936	36.81	3,710	15.78	4.45	298	230	450
UH-1Y	Utility	11.35	5,370	8,390	49.90	3,092	14.63	4.44	366	250	686
AH-1Z	Combat	11.28	5,580	8,392	49.90	3,446	14.60	4.37	411	296	686
CH-47D	Cargo	20.20	10,151	22,680	47.00	7,500	18.60	5.70	298	256	741
AH-64	Attack	15.20	5,165	9,525	62.10	3,600	14.63	4.66	365	265	407
EC-145	Utility	6.37	1,804	3,585	37.70	1,540	11.00	3.96	268	241	680
AS-532UB	Utility	14.12	4,330	9,000	48.90	3,754	15.60	4.80	278	239	573
UH-60L	Utility	11.51	5,224	10,660	47.20	3,780	16.40	5.18	294	266	584
UH-72A LAKOTA	Utility	6.06	1,792	3,585	37.70	1,476	11.00	3.96	268	241	685

C. CONSTRUCTION OF CERS BY TRADITIONAL (OLS) METHODS

Both linear and power regressions were carried out, and from these regressions, the unit cost of the KUH was estimated.

1. Selection of Cost Driver. Regressions Were Carried Out For the Following Circumstances

a. Cost vs 1 Variable

The dependent variable is Average unit cost and the single independent variable is one of the nine cost drivers in order to evaluate the performance of eight types of helicopter.

b. Cost vs 2 Variables

The dependent variable is Average unit cost and the two independent variables are the combinations of the cost drivers. Such a model will show more specific relationships between the average unit cost and variables. There are 36 two-variable combinations to evaluate the performance of eight types of helicopter.

2. Methodology

Two means of regression, Linear and Power, were used to find the cost estimating models.

a. Linear Regression

The linear Models are expressed by the equations below:

- One dependent variable and one independent variable:

$$\boxed{\text{Cost} = A + B^*(\text{Variable 1})}$$

- One dependent variable and two independent variables:

$$\boxed{\text{Cost} = A + B^*(\text{Variable 1}) + C^*(\text{Variable 2})}$$

b. Power Regression Model

To model non-linear relationships with OLS regression, the data must first be transformed in a way that makes the relationship linear. All the steps for linear regression may then be performed on the transformed data.

$$y = A^*X^B \quad \Longleftrightarrow \quad \ln y = \ln A + B^* \ln X$$

The power regression models are expressed as follows:

- One dependent variable and one independent variable:

$$\boxed{\text{Cost} = A^*(\text{Variable 1})^B}$$

- One dependent variable and two independent variables:

$$\boxed{\text{Cost} = A^*(\text{Variable 1})^B {}^*(\text{Variable 2})^C}$$

c. Criteria of Evaluation

Using the OLS method, 90 CERs, 18 one-variable CERs and 72 two-variable CERs, were developed.

The statistical significance of these 90 CERs, was assessed, using the tests in Table 2.

Table 2. Criteria of Evaluation

R-square	F-Significance	P-value
≥ 0.7	≤ 0.1	≤ 0.1

(1) R-Square. This represents the proportion of total variation around $\overline{Y}$ (average cost) explained by the regression model. The larger, the better.

(2) F-Significance. This is a statistical test that compares the fit of the models to the fit of a model with only the parameter. A smaller value indicates a greater improvement.

(3) P-Value. This measures the improvement in the model where a single prediction is included. In the case of one independent variable, this will be identical to the F significance above. Again, a smaller value indicates a greater improvement.[8]

3. Results of Regression

As a result of the filtering for statistical significance described above, the 90 cases were reduced to 22 cases which satisfied the evaluation criteria. These 22 cases are displayed in Appendix II. Additionally, some of these results are displayed in the following tables, in which the regressions that passed all the evaluation criteria are highlighted.

Reviewing the 22 cases, we found that the variables, Dimension, Power Plant, Weight and Range, are the important factors in estimating cost. However, the variable, Speed, was less significant for estimating costs.

a. Linear Regression with One Variable

First, the one-variable linear regressions with average unit cost and nine cost driver factors were executed. Among nine variables, four variables, Max Taking-off, SHP, Height and Empty weight, met the criteria of the evaluations. The results are shown in Table 3.

[8] Douglas C. Montgomery, Elizabeth A. Peck, and G.Geoffrey Vining, Introduction to Linear Regression Analysis, (Hoboken, New Jersey: Wiley-Interscience, 2006), 26, 44.

Table 3. The Results of 1 Variable Linear Regression

Dependent variable	Independent variable	Linear Regression with 1 variable				
		P-value	Significance F	R Square	Equation	Estimation
Average Unit Cost	Max Taking-Off	0.0014	0.001	0.8374	$y = 0.0007x_1 + 5.2583$	11.515
	Max disc loading	0.1002	0.1002	0.3859	$y = 0.3719x_1 - 5.673$	8.017
	SHP	0.0005	0.0005	0.8884	$y = 0.0023x_1 + 3.7471$	12.280
	Main Rotor	0.0163	0.0163	0.6456	$y = 1.6421x_1 - 11.894$	14.018
	Height	0.0037	0.0037	0.7788	$y = 6.8692x_1 - 19.82$	10.752
	Max speed	0.5957	0.5957	0.0497	$y = 0.019x_1 + 5.9695$	11.632
	Cruising speed	0.6003	0.6003	0.0485	$y = 0.0534x_1 - 1.7029$	10.579
	Max Range (km)	0.6870	0.6870	0.0290	$y = -0.0074x_1 + 16.681$	13.351
	Empty Weight	0.0015	0.0015	0.8363	$y = 0.0016x_1 + 4.0405$	11.917

(a) Range of linear estimation: 10.75 ~ 12.28 ($MFY08)

(b) Average of linear cost estimation: 11.616 ($MFY08)

(c) Standard Deviation: 0.6557

b. Power Regression with One Variable

Next, one variable power regressions were carried out with average unit cost and one of nine cost driver factors. Among nine variables, five variables, Max Taking-off, SHP, Main rotor, Height and Empty weight, met the criteria of the evaluations. The results are shown in Table 4.

Table 4. The Results of 1 Variable Power Regression

Y	Independent variable	Power Regression with 1 variable				
		P-value	Significance F	R Square	Equation	Estimation
Average Unit Cost	Max Taking-Off	0.0002	0.0002	0.9125	$y = 0.0297\ X_1^{0.659}$	11.928
	Max disc Loading	0.0287	0.0287	0.5775	$y = 0.0065X_1^{1.9362}$	6.997
	SHP	0.0001	0.0001	0.9372	$y = 0.0245\ X_1^{0.7609}$	12.738
	Main Rotor	0.0031	0.0031	0.7363	$y = 0.0401\ X_1^{2.1137}$	13.664
	Height	0.0041	0.0041	0.7719	$y = 0.1377\ X_1^{2.8812}$	10.162
	Max speed	0.3690	0.3690	0.1358	$y = 0.0559\ X_1^{0.9215}$	10.651
	Cruising speed	0.4329	0.4329	0.1053	$y = 0.0004\ X_1^{1.8592}$	9.840
	Max Range	0.5032	0.5032	0.0779	$y = 531.51\ X_1^{(-0.6)}$	13.601
	Empty Weight	0.0006	0.0006	0.8802	$y = 0.0468\ X_1^{0.6547}$	12.233

(a) Range of power regression: 10.16 ~ 13.66 ($MFY08)

(b) Average of linear cost estimation: 12.1451($MFY08)

(c) Standard Deviation: 1.2890

c. Linear Regression with Two Variables

In order to find more specific cost drivers, two variable linear regressions were examined with average unit costs and 36 combinations of two variables from nine cost driver factors. The results appear in Table 5 and Appendix II. A.

Table 5. The Results of Two-Variable Linear Regression

Y	Independent Variable		Linear Regression with 2 variables						
	X_1	X_2	P-value				Significance F	R Square	Estimation
Unit Cost	Max Taking-Off	Max disc loading	X_1	0.0004	X_2	0.0135	0.0004	0.9571	9.324
			Equation	$y = -4.2752 + 0.000621X_1 + 0.218676X_2$					
		Max Range	X_1	0.0004	X_2	0.0370	0.0010	0.9373	14.114
			Equation	$y = 13.670328 + 0.000751X_1 - 0.013928\ X_2$					
	Max disc loading	SHP	X_1	0.0112	X_2	0.0001	0.0001	0.9726	10.381
			Equation	$y = -4.133102 + 0.187267X_1 + 0.002054\ X_2$					
		Height	X_1	0.1003	X_2	0.0065	0.0052	0.8778	8.749
			Equation	$y = -24.908503 + 0.203015\ X_1 + 5.884112\ X_2$					
	SHP	Max Range	X_1	0.0001	X_2	0.0183	0.0002	0.9669	14.678
			Equation	$y = 11.204742 + 0.002426\ X_1 - 0.012283\ X_2$					
	Max Range	Empty Weight	X_1	0.0507	X_2	0.0005	0.0013	0.9291	14.423
			Equation	$y = 12.10339 - 0.0134\ X_1 + 0.001696\ X_2$					

(a) Range of linear regression: 8.75 ~ 14.48 ($MFY08)

(b) Average of linear cost estimation: 11.945 ($MFY08)

(c) Standard Deviation: 2.7512

d. Power Regression with Two Variables

In order to find more specific cost drivers and to fit the non-linear to linear, two-variable power regressions were examined with average unit costs and 36 combinations of two variables from nine cost driver factors, which the results displayed in Table 6 and Appendix II.B.

Table 6. The Results of Two-Variable Power Regression

Y	Independent Variable X_1	Independent Variable X_2	P-value X_1	P-value X_2	Significance F	R Square	Estimation	Equation
Average Unit Cost	Max Taking-Off	Max disc loading	0.0008	0.0505	0.0003	0.9622	9.896	$y=0.005459*X_1^{0.540361}*X_2^{0.717311}$
		Max Range	0.0002	0.0746	0.0004	0.9564	13.787	$y=0.599107*X_1^{0.648372}*X_2^{(-0.452229)}$
	Max disc loading	SHP	0.0398	0.0003	0.0001	0.9751	10.667	$y=0.005687* X_1^{(0.637706})*X_2^{(0.637240)}$
		Main Rotor	0.1109	0.0039	0.0013	0.9308	10.988	$y=0.006873*X_1^{0.738957}*X_2^{1.708173}$
		Height	0.0216	0.0043	0.0014	0.9280	7.872	$y=0.004887*X_1^{1.138692}*X_2^{2.196166}$
	SHP	Max Range	0.00004	0.0368	0.0001	0.9758	14.561	$y=0.417159*X_1^{0.747558}*X_2^{(-0.424173)}$
	Height	Max speed	0.0024	0.0819	0.0047	0.8827	9.729	$y=0.001226*X_1^{2.836414}*X_2^{0.832812}$

(a) Range of power regression: 7.87 ~ 14.56 ($MFY08)

(b) Average of linear cost estimation: 11.0713 ($MFY08)

(c) Standard Deviation: 2.3502

e. Analysis of the Results for Traditional OLS

(1) Comparison of average estimation of the KUH. An one variable power regression estimating cost produced the highest and a two-variable linear regression model cost estimating produced the second highest value.

The distribution of average cost is in the range of 11.07~12.15($MFY08) as shown in Table 7.

Table 7. The Average of Estimation

Type	Estimation	
	Linear	Power
1variable	11.62	12.15
2variable	11.94	11.07

(2) Stability of cost estimation of the KUH. By checking the average and Max-Min estimation, it can be seen that estimates from the one-variable linear regression model is distributed narrowly, providing confidence in the estimates.

But, the stability of data must be confirmed by testing the standard deviations of the predictions, where smaller standard deviations are better than larger values.

A one-variable linear regression model has the smallest standard deviation and is the most attractive model as shown in Table 8.

Table 8. The Standard Deviation of Estimation

Type	Standard Deviation	
	Linear	Power
1 variable	0.6557	1.2890
2variable	2.7512	2.3502

(3) Confidence interval for the cost estimation of the KUH. We constructed 95 percent confidence intervals for the predictions, shown in Table 9. T-statistics are used because the sample size was less than 30.

This also shows that a one-variable linear model has the narrowest 95 percent confidence level. The one-variable linear model appears the most promising model.

Table 9. The Confidence Interval with 95 Percent Confidence Level

1 variable	Linear	Power	2 variables	Linear	Power
Sample Size	4	5	Sample Size	6	7
95% Lower	11.603	12.124	95% Lower	11.904	11.040
95% Higher	11.628	12.167	95% Higher	11.985	11.103
Difference	0.025	0.043	Difference	0.081	0.063

D. CONSTRUCTION OF THE KUH CERS BY ADAPTIVE CER

This method is similar to the approach used in the previous paragraph. For OLS, one- and two-variable linear regressions were used, and a one- and two-variable regression method. Then, the average unit cost of the KUH was estimated.

The basic procedures for applying the adaptive CERs are the same as the traditional cost-estimating method from data collection to analysis of regressions. But, at this stage, the individual cost driver factors need to be transformed by applying weights to each variable.[9] Using weighted data, the procedures were repeated.

1. Methodology for Selecting Weights

Before applying the weighted least square (WLS) method, it is important to determine how much weight is assigned to an individual helicopter. The transformed data is displayed in Appendix II.C. The way of selecting weights used is:

1. Remove the unnecessary variable. Cruising speed was removed from cost drivers because the cruising speed was not a significant factor in estimating costs.
2. Compare the similarity between the KUH and other helicopters using the eight cost drivers. This computation of "initial weight value" is displayed in the equation below.

$$AUH1Y = \left| \sum\left(\frac{UH1Y\ data}{KUH\ data}\right) - 8 \right| / 8$$

[9] Book, Broder and Feldman, "Statistical Foundations of Adaptive Cost-Estimating Relationships," 5–6.

3. The absolute value of the initial weight value was mapped into the scale from 1 to 10, as indicated in the Table 10.

Table 10. Initial Weight

Interval	0~0.1	0.1~0.2	0.2~0.3	0.3~0.4	0.4~0.5	0.5~0.6	0.6~0.7	0.7~0.8	0.8~0.9	0.9~1.0
Initial weight	10	9	8	7	6	5	4	3	2	1

4. To compute the "modified weight" from the initial weight, we multiply the initial weight by a penalty, which depends on the purpose of the helicopter, as shown Table 11.

Table 11. Penalty by Purpose

Purpose of helicopter	Utility medium	Utility	Other
Penalty	1	0.9	0.8

5. To normalize the weight, each modified weight is divided by the sum of modified weight.

$$\textbf{Normalized weight} = \frac{Modified\,weight}{\sum modified\ weight}$$

6. Multiply each X and Y by the square root of normalized weight assigned to each helicopter in Table 12.

Table 12. Example of Selecting Weight for UH-1Y

Name	Type	Sum of ratio	A_K	Initial weight	Modified weight	Normalized	Sqrt(weight)
UH-1Y	Medium Utility	8.8962	0.1120	9	1	0.15	0.3873

2. Selection of Cost Drivers

a. Cost vs. 1 Variable

The dependent variable is the weighted average unit costs of the eight helicopters in the database. The weighted independent variables are one of the five cost drivers.

b. Cost vs. 2 Variables

The dependent variable is the weighted average unit costs of the eight helicopters in the database. The weighted independent variables are two of the cost drivers, chosen from the five cost drivers.

3. Methodology

Two ways of regression, Linear and Power, were used to develop the cost estimating models. These are described below. This is the same method which was executed in OLS method.

a. Linear Regression

The linear Models are expressed by the equations below: There are four equations using one variable, and there are six equations using two variables.

- one dependent variable and one independent variable:

$$\boxed{\text{Cost} = A + B\text{*(Variable 1)}}$$

- one dependent variable and two independent variables:

$$\boxed{\text{Cost} = A + B\text{*(Variable 1)} + C\text{*(Variable 2)}}$$

b. Power Regression Model

To model non-linear relationships with WLS regression, the data must first be transformed in a way that makes the relationship linear. All the steps for linear regression may then be performed on the transformed data.

$$y = A{*}X^{b} \Longleftrightarrow \qquad \ln y = \ln a + b{*} \ln x$$

The power regression models are expressed as follows:

- One dependent variable and one independent variable:

$$\boxed{\text{Cost} = A\text{*(Variable 1)}^{B}}$$

- One dependent variable and two independent variables:

$$\text{Cost} = A^{*}(\text{Variable 1})^{B} {}^{*}(\text{Variable 2})^{C}$$

By WLS and power regression, five cases of one variable and seven cases of two variables cost estimating models were developed.

c. *Criteria of Evaluation*

At the same time, the regression results had to be examined to know how much they were fit for the real data. And, the level of independence of variables to each other needed to be checked to obtain more appropriate models by following Table 13.

Table 13. Criteria of Evaluation

R-square	F-Significance	P-value
≥ 0.7	≤ 0.1	≤ 0.1

4. Results of Regression by Weighted Variables

Using OLS and power regression, eight cases of one-variable and 13 cases of two -variable cost estimating models were constructed.

Power Plant and related performance proved to be more important factors to estimate cost. However, Speed, dimension and range variables were less significant for affecting the relation of unit cost and each factor.

a. *Linear Regression with 1 Weighted Variable*

There is one weighted variable, SHP for power plant that satisfies the criteria of evaluation in Table 14.

Table 14. The Result of Linear Regression with one Weighted Variable

Y	Independent Variable	Linear regression with one weighted variable				
		R Square	P-value	Significance F	Equation	Estimate
Unit Cost	Max Taking-Off	0.6801	0.0118	0.0001	$y = 0.0008\ x_1 + 1.665$	8.814
	SHP	0.7827	0.0035	0.0000	$y = 0.0026\ x_1 + 0.9884$	10.634
	Height	0.3865	0.0998	0.0004	$y = 3.082\ x_1 - 0.8646$	12.850
	Empty Weight	0.6599	0.0143	0.0001	$y = 0.0016\ x_1 + 1.3986$	9.275

(a) Range of power regression: 10.63 ($MFY08)

(b) Average of linear cost estimation: 10.63 ($MFY08)

(c) Standard Deviation is not determined

b. Power Regression with One Weighted Variable

After that, one-variable power regression was carried out with weighted average unit cost and one of five weighted cost driver factors. Among five variables, three variables, Max Taking-off, SHP, and Empty weight, met the criteria of evaluations. The results are in a Table 15.

Table 15. The Result of Power Regression with One Weighted Variable

Y	Independent Variable	Power regression with one weighted variable				
		R Square	P-value	Significance F	Equation	Estimate
Average Unit Cost	Max Taking-Off	0.8426	0.0058	0.0014	$y = 0.0212 X_1^{0.6557}$	8.262
	SHP	0.8976	0.0003	0.0004	$y = 0.0178\ X_1^{0.7701}$	9.982
	Main rotor	0.6344	0.0180	0.0165	$y = 0.3731\ X_1^{1.4604}$	20.967
	Height	0.3818	0.1117	0.1029	$y = 1.9793\ X_1^{1.4444}$	17.100
	Empty Weight	0.8341	0.0080	0.0016	$y = 0.036\ X_1^{0.6407}$	8.354

(a) Range of power regression: 8.26~9.98 ($MFY08)

(b) Average of linear cost estimation: 8.87 ($MFY08)

(c) Standard Deviation: 0.9670

c. Linear Regression with Two Weighted Variables

In order to find more specific cost drivers, two-variable linear regressions were examined with weighted average unit cost and six combinations of two variables among eight cost driver factors. As a result, two cost-estimating models were derived.

Table 16. The Result of Linear Regression Two Weighted Variables

Y	Independent Variable		Linear regression with two weighted variables						
	X_1	X_2	P-value				Significance F	R Square	Estimate
Average Unit Cost	Max Taking-Off	Max disc loading	X_1	0.0028	X_2	0.0095	0.0039	0.9167	11.835
			Equation	$y = -1.137612 + 0.000696X_1 + 0.183470X_2$					
		Max Range	X_1	0.0472	X_2	0.8270	0.0905	0.6422	7.947
			Equation	y= 2.229719 + 0.000750X1 - 0.002189X2					
	Max disc loading	SHP	X_1	0.0117	X_2	0.0009	0.0016	0.9457	12.766
			Equation	$y = -0.992533 + 0.145093X_1 + 0.002269X_2$					
		Height	X_1	0.4507	X_2	0.5661	0.2404	0.4589	11.304
			Equation	$y = -0.733044 + 0.141554X_1 + 1.534076X_2$					
	SHP	Max Range	X_1	0.0117	X_2	0.7339	0.0282	0.7882	9.977
			Equation	$y = 1.649922 + 0.0025555\ X_1 - 0.002564X_2$					
	Max Range	Empty Weight	X_1	0.4989	X_2	0.0313	0.0642	0.6925	1.952
			Equation	$y = -2.886468 - 0.006052X_1 + 0.001536X_2$					

(a) Range of power regression: 11.83~12.76 ($MFY08)

(b) Average of linear cost estimation: 12.30 ($MFY08)

(c) Standard Deviation: 0.6582

d. Power Regression with Two Variables

In order to find more specific cost drivers, two-variable power regressions were examined with weighted average unit cost and seven combinations of two-variable among eight cost driver factors.

However, while the Excel regression tool was executed with two weighted variables, it turned out different results following the option whether checking the “constant is zero or not” as shown in Table 17. So it appears difficult to evaluate if the derived values are feasible or not.

Even though obtaining results was attempted the result of power regression with 2 weighted variables has been excluded.

Table 17. The Result of Two Weighted Variables Power Regression

	Data		Power Regression with two weighted variables							
	X_1	X_2	P-value				Significance F	R Square	Estimate	
Unit Cost vs.	Max Taking-Off	Max disc loading	X_1	0.3415	X_2	0.9891	0.0014	0.9506	4.8096	11.7196
			Equation	$y=X_1^{0.16999}*X_2^{0.006673}$						
			Equation	$y=0.008711*X_1^{0.581314}*X_2^{0.531295}$						
		Max Range	X_1	0.0022	X_2	0.0141	0.0144	0.8451	4.8850	7.5755
			Equation	$y=X_1^{0.513960}*X_2^{(-0.505750)}$						
			Equation	$y=0.039933*X_1^{0.641627}*X_2^{(-0.096895)}$						
	Max disc loading	SHP	X_1	0.9210	X_2	0.3082	0.0007	0.9646	4.9195	12.8004
			Equation	$y=X_1^{(-0.050217)}*X_2^{0.215881}$						
			Equation	$y=0.009652*X_1^{0.429326}*X_2^{0.686478}$						
		Main Rotor	X_1	0.3007	X_2	0.0646	0.0926	0.6387	18.7273	21.9242
			Equation	$y=X_1^{(-0.493376)}*X_2^{1.706931}$						
			Equation	$y=0.460299*X_1^{(-0.163975)}*X_2^{1.614766}$						
		Height	X_1	0.3170	X_2	0.7703	0.0062	0.8692	14.6989	14.3167
			Equation	$y=X_1^{0.310966}*X_2^{1.049298}$						
			Equation	$y=0.795303*X_1^{0.410044}*X_2^{0.945762}$						
	SHP	Max Range	X_1	0.0008	X_2	0.0055	0.0054	0.9029	5.5036	8.7857
			Equation	$y=X_1^{0.609733}*X_2^{(-0.541124)}$						
			Equation	$y=0.043340*X_1^{0.749378}*X_2^{(-0.138671)}$						
	Height	Max speed	X_1	0.1522	X_2	0.1220	0.3298	0.3847	23.6934	52.3946
			Equation	$y=X_1^{1.351780}*X_2^{0.154325}$						
			Equation	$y=3.027244*X_1^{1.531240}*X_2^{0.099201}$						

e. Analysis of the Result

The author developed 22 significant OLS models. When the variables from these models were recast as WLS models, only six survived the fitness criteria.

(1) Comparison of average estimation of the KUH. The results in the WLS case differ from the results in the OLS case. In the WLS case, two-weighted variables linear regression estimating cost is the highest and one-weighted-variable regression model is the lowest cost estimation, which is indicated in Table 18.

Table 18. The Average of Estimation by WLS

Type	Average of Estimation by W.L.S	
	Linear	Power
One variable	10.63	8.87
Two variables	12.30	N/A

(2) Stability of cost estimation of the KUH. By checking the average and Max-Min estimation, it can be recognized that the one-variable linear regression model estimates are distributed narrowly, providing confidence in the estimates.

But, the stability of the data should be confirmed by testing the standard deviations of data. The smaller the value is, the better the stability of the estimation.

One-variable linear regression model has the smallest standard deviation and it is the most attractive model.

In this case, both models have small standard deviation. Both of them are attractive models in Table 19.

Table 19. The Standard Deviation by W.L.S

Type	Standard Deviation by W.L.S	
	Linear	Power
1 variable	N/A	0.967
2variable	0.6582	N/A

(3) Confidence interval for the cost estimation of the KUH. With 95 percent confidence level, confidence intervals are measured as Table 20. T-statistics is used because the sample size is less than 30.

Only two types of models were tested based upon the significance of our results. It shows that the weighted, two-variable linear model has the narrowest interval with 95 percent confidence level. This Adaptive CER appears the most confident prediction of cost.

Table 20. The Confidence Interval with 95 Percent Confidence Level by W.L.S

1 variable	Linear	Power	2 variables	Linear	Power
Sample Size	1	3	Sample Size	2	7
95% Lower	N/A	8.8419	95% Lower	12.2750	N/A
95% Higher	N/A	8.8902	95% Higher	12.3267	N/A
Difference	N/A	0.483	Difference	0.0517	N/A

E. COMPARISON AND EVALUATION

The results derived were compared and are displayed in the Table 21.

It was found that the error (Standard Deviation) term for WLS is less than the standard deviation for OLS, which in fact is the objective of doing WLS. Overall, WLS models have standard deviations that are similar to or smaller than OLS models.

At the same time, the difference of average should be considered. Most cases show the gap within 10 percent of variation. But, the one variable power regression model has a gap of 3.28 $MFY08. It may be caused by the lack of comparison data.

While any of these models is acceptable, it is author’s opinion that the two-variable linear WLS model is particularly attractive for use in estimating the unit cost of the KUH.

Table 21. The Comparison of Estimation from OLS and WLS

Method		Number of Variable	Number of models	Estimates of KUH	
				Average	Standard Deviation
OLS	Linear	1 variable	4	11.62	0.66
		2 variables	6	11.94	2.75
	Power	1 variable	5	12.15	1.29
		2 variables	7	11.07	2.35
WLS	Linear	1 variable	1	10.63	N/A
		2 variables	2	12.30	0.66
	Power	1 variable	3	8.87	0.97

IV. CONCLUSION AND RECOMMENDATIONS

A. CONCLUSION

Cost estimation and analysis is very important for government acquisition programs for many reasons: to support funding decisions, to evaluate resource requirement at key decision points, and to develop performance measurement baselines.

ROKA (Republic of Korea Army) made plan to replace the old version of helicopters to improve capability for operational requirements and has carried out the KUH (Korea Utility Helicopter) program from KHP (Korea Helicopter Program) since 2005. After success of KUH, ROKA will continue to develop the KAH (Korea Attack Helicopter) based on KUH.

The author attempted to develop the CER for the KUH using traditional OLS and WLS of the adaptive CER method and implemented 8 kinds of models to find more feasible relationship. Ninety estimates from OLS and 22 estimates from WLS were analyzed.

By examining various conditions and methods, the author of the thesis found that adaptive CER methodology can provide a more stable prediction of costs for the KUH.

A prototype of KUH has already been produced and is undergoing testing. If it passes the testing phase, the program will transition into the manufacturing phase. At the same time, KHP will start on the foundation of KUH, where it will also need to estimate the cost. By applying the adaptive CER method to KHP with more abundant data, we will have a better basis for CER development and accurate cost estimates.

B RECOMMENDATIONS FOR FUTURE WORK

Eight kinds of specific methods (linear/power; 1- and 2-variable; OLS and WLS) with nine independent variables at the helicopter-system level were carried out. These methods provided a varied set of cost estimates for the KUH.

However, a further range of research is needed to derive more accurate cost estimates. This future research should include:

- More data gathered and evaluated for this thesis, only 9 cost driver factors were collected due to the limits of data collection. If more data of performance and specifications were used, over or under cost estimation would be reduced.
- Second, the more models tested, the better cost estimating relationships will be derived. Finally, while designing and researching the KUH, additional cost data for subsystems of the KUH could be obtained. Models should be expanded from the system level to the level of subsystems and main components such as Work Breakdown System (WBS) including armament and avionics.

LIST OF REFERENCES

한국형 헬기사업(Hankookhyung Helgisaup), Wikipedia, http://ko.wikipedia.org/wiki/%ED%95%9C%EA%B5%AD%ED%98%95_%ED%97%AC%EA%B8%B0_%EC%82%AC%EC%97%85 (Accessed July 28, 2009).

AH-1Z Viper, Wikipedia, http://en.wikipedia.org/wiki/AH-1Z

Bell Helicopter Textron Inc., http://www.bellhelicopter.com/en/aircraft/military/pdf/UH1Y_PG_3-06_web.pdf

Bell Helicopter Textron Inc., http://www.bellhelicopter.com/en/aircraft/military/pdf/AH1Z_PG_3-06_web.pdf

Boeing 114 (CH-47) Chinook - CH-47F Improved Cargo Helicopter (ICH) Upgrade, http://search.janes.com/Search/documentView.do?docId=/content1/janesdata/yb/jau/jau_1465.htm@current&pageSelected=allJanes&keyword=CH-47&backPath=http://search.janes.com/Search&Prod_Name=JAU&#toclink-j0010100032253 (Accessed August 2, 2009).

BOEING AH-64 APACHE, Jane’s All the World’s Aircraft, 11-Feb-2009, http://search.janes.com/Search/documentView.do?docId=/content1/janesdata/yb/jawa/jawa1193.htm@current&pageSelected=allJanes&keyword=AH-64&backPath=http://search.janes.com/Search&Prod_Name=JAWA&#toclink-j0010120126531 (Accessed Aug 15, 2009).

BOEING AH-64 APACHE, Jane’s All the World’s Aircraft, 13-Oct- 2009, http://www.janes.com/defence/air_forces/news/jawa/jawa001013_1_n.shtml (Accessed Oct 19, 2009).

CH-47 Chinook, Wikipedia, http://en.wikipedia.org/wiki/CH-47 (Accessed August 2, 2009).

Cougar, AS-532, Eurocopter http://www.eurocopter.com/site/en/ref/Characteristics_160.html (Accessed Aug 10, 2009).

EADS (Eurocopter) UH-72A Lakota, Jane’s All the World’s Aircraft, 11-May-2009, http://search.janes.com/Search/documentView.do?docId=/content1/janesdata/yb/jawa/jawaa430.htm@current&pageSelected=allJanes&keyword=uh-72&backPath=http://search.janes.com/Search&Prod_Name=JAWA& (Accessed Aug 5, 2009).

EC 145, Eurocopter, http://www.eurocopter.com/site/en/ref/Characteristics_93.html (Accessed Aug 7, 2009).

Eurocopter AS 332 Super Puma and AS 532 Cougar, Jane's Helicopter Markets and Systems, 03-Apr-2009, http://search.janes.com/Search/documentView.do?docId=/content1/janesdata/binder/jhms/jhms5004.htm@current&pageSelected=allJanes&keyword=as%20532&backPath=http://search.janes.com/Search&Prod_Name=JHMS& (Accessed Aug 10, 2009).

Eurocopter AS 532, Wikipedia, http://en.wikipedia.org/wiki/AS-532 (Accessed Aug 10, 2009).

Eurocopter EC 145, Jane's Helicopter Markets and Systems, 23-Oct-2009, http://search.janes.com/Search/documentView.do?docId=/content1/janesdata/binder/jhms/jhms9168.htm@current&pageSelected=allJanes&keyword=ec-145&backPath=http://search.janes.com/Search&Prod_Name=JHMS&#toclink-j0010200018250 (Accessed Aug 7 2009). http://www.boeing.com/rotorcraft/military/ah64d/docs/AH-64D_overview.pdf (Accessed Aug 15, 2009), http://www.boeing.com/rotorcraft/military/ch47d/docs/CH-47D_overview.pdf (Accessed August 3, 2009).

KAI Surion, *Jane's All the World's Aircraft*, http://search.janes.com/Search/documentView.do?docId=/content1/janesdata/yb/jawa/jawaa333.htm@current&pageSelected=allJanes&keyword=kuh&backPath=http://search.janes.com/Search&Prod_Name=JAWA& (Accessed November 25, 2009).

KAI SURION, Wikipedia, http://en.wikipedia.org/wiki/KAI_Surion (Accessed July 28, 2009).

Paul E. Eden and Soph Moeng, *Modern Military Aircraft Anatomy Technical Drawings of 118 Aircraft 1945 to the Present Day*, New York: MetroBooks, 2002 .

Sikorsky S-70A, Jane's All the World's Aircraft, 11-Feb-2009, http://www8.janes.com/Search/documentView.do?docId=/content1/janesdata/yb/jawa/jawa1457.htm@current&pageSelected=allJanes&backPath=http://search.janes.com/Search&Prod_Name=JAWA&keyword=   (Accessed Aug 11, 2009).

Stephen A. Book, Melvin A. Border and Daniel I. Feldman, "Statistical Foundations of Adaptive Cost-Estimating Relationships," SCEA (Society of Cost Estimating and Analysis)-ISPA (International Society of Parametric Analysts) Joint Annual Conference & Training Workshop, June 24–27, 2008.

Stephen A. Book, Melvin A. Border and Daniel I. Feldman, “Adaptive Cost-Estimating Relationships”, SCEA(Society of Cost Estimating and Analysis)-ISPA(International Society of Parametric Analysts) Joint Annual Conference & Training Workshop, June 24-27, 2008), 2–3.

Sungjin Kang, Gyumyung Choi, Jongbok Jung, and Seungsoo Kim, “KMH Cost Analysis Report,” Korea National Defense University(KNDU) Report for Korea DAPA, Dec 2005.

UH-1Y Venom, Wikipedia, http://en.wikipedia.org/wiki/UH-1Y_Venom

UH-60 Black Hawk, Wikipedia, http://en.wikipedia.org/wiki/UH-60_Black_Hawk#Specifications_.28UH-60L.29 (Accessed Aug 11, 2009).

UH-72 Lakota, Wikipedia, http://en.wikipedia.org/wiki/UH-72_Lakota (Accessed Aug 7, 2009).

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A. STATISTICAL FOUNDATIONS OF ADAPTIVE COST-ESTIMATING RELATIONSHIPS

This paper is in the public domain and is available from the Society of Cost estimating and Analysis 2008 conference proceedings.

Statistical foundations of adaptive cost-estimating relationships

Stephen A. Book, MCR LLC
Melvin A. Broder, The Aerospace Corporation
Daniel I. Feldman, MCR LLC

Abstract

Traditional development of cost-estimating relationships (CERs) has been based on "full" data sets consisting of all available cost and technical data associated with a particular class of products of interest, e.g., components, subsystems or entire systems of satellites, ground systems, etc. In this paper, we review an extension of the concept of "analogy estimating" to parametric estimating, namely the concept of "adaptive" CERs—CERs that are based on specific knowledge of individual data points that may be more relevant to a particular estimating problem than would the full data set. The goal of adaptive CER development is to be able to apply CERs that have smaller estimating error and narrower prediction bounds. Several examples of adaptive CERs were provided in a paper (Reference 2) presented by the first two authors to the May 2008 SSCAG Meeting in Noordwijk, Holland, and the July 2008 ISPA/SCEA Conference in Industry Hills CA.

This paper focuses on statistical foundations of the derivation of adaptive CERs, namely the method of weighted least-squares (WLS) regression. Ordinary least-squares (OLS) regression has been traditionally applied to historical-cost data in order to derive additive-error CERs valid over an entire data range, subject to the requirement that all data points are weighted equally and have residuals that are distributed according to a common normal distribution. The idea behind adaptive CERs, however, is that data points should be "deweighted" based on some function of their distance from the point at which an estimate is to be made, i.e., each historical data point should be assigned a

"weight" that reflects its importance to the particular estimation that is to be made using the derived CER. This presentation describes technical details of the WLS derivation process, resulting quality metrics, and the roles it plays in adaptive-CER development.

Introduction

Weighted least-squares (WLS) regression is the statistical technique applied in Reference 1 to develop adaptive CERs. WLS regression is a straightforward extension of classical ordinary least-squares (OLS) regression, which is the 18th Century curve-fitting technique commonly taught in elementary statistics courses.

OLS regression "best" fits a straight line $\boldsymbol{y = a + bx}$ to a set of ordered pairs $\boldsymbol{(x_k, y_k)}$, $\boldsymbol{1 \le k \le n}$, of data points in two-dimensional Euclidean space. We will get to the OLS definition of "best" momentarily. Procedures based on OLS philosophy and mathematical principles can extend OLS regression to the case of curved lines, primarily logarithmic, as well as a multidimensional context. However, for our purposes of deriving adaptive CERs, the linear two-dimensional context suffices.

Suppose we have $\boldsymbol{n}$ data points such as those in Table 22, labeled $\boldsymbol{(x_1, y_1)}$, $\boldsymbol{(x_2, y_2)}$, …, $\boldsymbol{(x_n, y_n)}$, where, for $\boldsymbol{1 \le k \le n}$, $\boldsymbol{y_k}$ is the actual cost associated with a program whose cost driver (perhaps weight, power, etc.) is $\boldsymbol{x_k}$. Were we to use the OLS regression line $\boldsymbol{y = a + bx}$ to predict the cost of the program in question, our cost estimate would have been $\boldsymbol{a + bx_k}$, rather than the actual cost $\boldsymbol{y_k}$. The equation $\boldsymbol{y = a + bx}$ is therefore called a "cost-estimating relationship" (CER).

Program	Cost-Driver Value x	Unit Cost y
A	156.12	51,367.22
B	179.40	5,885.00
C	180.30	7,060.00
D	217.50	139,483.12
E	419.14	3,386.00
F	437.09	6,738.00
G	440.93	6,812.00
H	494.45	3,291.34
I	789.90	5,723.14
J	826.10	10,992.00
K	864.30	11,590.00
L	869.30	15,973.00
M	976.50	7,970.67
N	1,355.80	9,524.10
O	1,360.90	35,927.22
P	1,463.21	11,238.73
Q	2,332.10	92,059.97
R	3,017.73	74,649.00
S	3,253.00	42,915.23

Table 22. Example of Historical Cost Data (19 Data Points)

The error in our estimate of the cost of any program is the difference $d_k = y_k - (a+bx_k) = y_k - a - bx_k$ between the actual cost y_k and the CER-estimated cost $a + bx_k$. The principle of least squares asserts that, in order to calculate the "best"-fitting straight line, we ought to choose the coefficients a and b, which determine the CER, so that the sum of squared differences (i.e., estimating errors)

$$f(a,b) = \sum_{k=1}^{n} d_k^2 = \sum_{k=1}^{n} (y_k - a - bx_k)^2$$

is as small as possible. By considering this problem as a two-dimensional minimization problem, we can take the partial derivatives of $f(a,b)$ with respect to a and b, respectively, set both partial derivatives equal to 0, and solve the resulting simultaneous equations for the two unknowns a and b. This process results in the following OLS explicit expressions for the slope b and the intercept a of the linear CER $y = a + bx$:

$$b = \frac{n\sum_{k=1}^{n} x_k y_k - \left(\sum_{k=1}^{n} x_k\right)\left(\sum_{k=1}^{n} y_k\right)}{n\sum_{k=1}^{n} x_k^2 - \left(\sum_{k=1}^{n} x_k\right)^2}$$

$$a = \frac{\sum_{k=1}^{n} y_k}{n} - b\frac{\sum_{k=1}^{n} x_k}{n}.$$

The above discussion summarizes what can be referred to as “naïve” regression. It is naïve, because a number of unstated assumptions that critically affect the nature of the CER and how it can be correctly applied are being made, often without the knowledge or concurrence of the cost analyst. The most important of these assumptions is that all n data points are and ought to be treated equally by the mathematical computations. An immediate unfortunate corollary is that extreme outlying data points, those far away from the bulk of the data and/or the cost-driver value at which the analyst wants to make an estimate, exert excessive influence on the location of the regression line and all estimates made using it.

What is it about OLS that requires us to consider each data point of equal merit? The answer to this question goes back to the early part of the 18th Century when it was mathematically derived from reasonable assumptions that estimation errors are well-modeled by the normal distribution. In fact, use of the word "normal" was introduced in the context of “the normal law of error” by Karl Pearson (1857-1936), a British scientist who was one of the founders of modern statistical theory. (It is said that Pearson later regretted his use of the word “normal,” coming to believe that its common usage biased less knowledgeable analysts against other statistical distributions, which they assumed to be “abnormal” in some sense.) The theory of regression assumes that the regression line is the truth and any departures from it, e.g., those in Figure 1 below, are errors. This means that the actual y values corresponding to any particular x value are normally distributed with mean equal to the number $a + bx$. Another way of looking at the OLS regression model is as $y_k = a + bx_k + \varepsilon_k$, where ε_k is a normally distributed random variable with mean 0 and standard deviation σ.

So far so good. The killer as far as CERs are concerned, though, is the OLS requirement that all normal distributions of $\boldsymbol{y}$ values (i.e., ε_k values), one for each $\boldsymbol{x}$ value, have the same standard deviation $\boldsymbol{\sigma}$. It is this requirement that forces OLS to consider all data points to be of equal merit. The requirement of equal $\boldsymbol{\sigma}$ values as a general rule, though, is highly questionable in the case of CERs, especially when the wide range of parameters on which CERs may be based is considered. Take a look at Figure 1. It seems clear that, for some technical reason as yet uninvestigated, cost is much more variable for cost-driver values near 300 than for other cost-driver levels. Why this happens should be studied in detail from the engineering point of view, but nevertheless we have to take account of it when estimating costs.

Figure 1 illustrates the data of Table 22, along with the OLS regression line that best fits the points in the least-squares sense. The dashed vertical lines in Figure 1 represent the distances $\boldsymbol{d_k}$ whose sum of squared values is to be minimized.

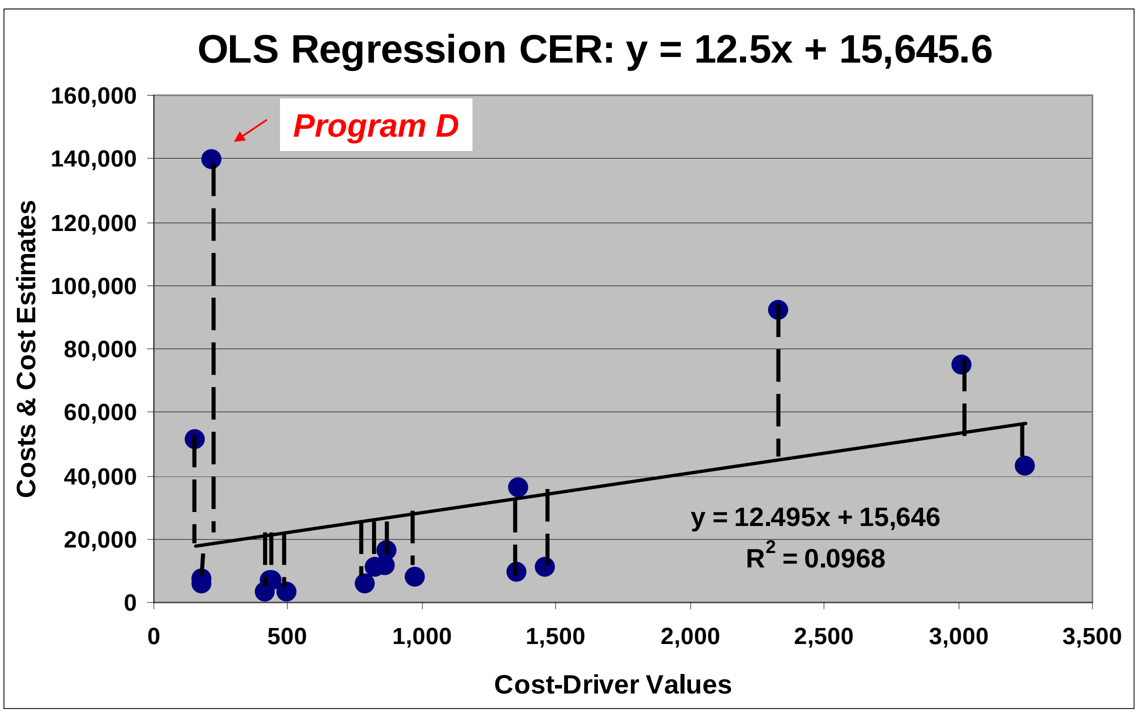

Figure 1. The Data Points of Table 1 and their OLS Regression Line

Consider the data point in Table 22 associated with Program D. From Figure 1, we see that this data point's d_k value will contribute the largest amount to the sum of squared estimating errors. In its attempt to minimize the sum of squared errors, the mathematics of OLS will take special pains to pull the regression line toward the Program D data point and thereby reduce the size of Program D's contribution to the total squared error. It is its very extremeness that gives the Program D data point its undue influence on the OLS regression line.

OLS CER Quality Metrics

Three quality metrics allow the cost analyst to assess the applicability of the CER to estimating problems involving the kinds of subsystems and/or components of which the supporting data base is comprised and the validity of estimates made using it. These three quality metrics are the following: (1) standard error of the estimate SEE; (2) bias B; and (3) R^2. We will discuss each of these in turn.

The standard error of the estimate SEE is an estimate of the σ value, which is the standard deviation of the normal distribution of $\varepsilon_k = y_k - a - bx_k$. Its expression is

$$SEE = \sqrt{\frac{\sum_{k=1}^{n}(y_k - a - bx_k)^2}{n-2}} = \sqrt{\frac{\sum_{k=1}^{n} y_k^2 - a\sum_{k=1}^{n} y_k - b\sum_{k=1}^{n} x_k y_k}{n-2}}\,.$$

In the OLS context, SEE is expressed in the same units as the costs and cost estimates, usually dollars. Because the coefficients of the OLS CER are calculated by minimizing the numerator under the square-root sign, the smaller the SEE turns out to be, the "better" the CER is. Choosing the denominator above as $n-2$ makes SEE an "unbiased" estimator of σ. If the denominator were simply n, SEE would be the "maximum-likelihood" estimator of σ, but not unbiased. "Unbiased" and "maximum likelihood" are statistical terms, for which we refer you to any advanced statistics text for further explanation.

The bias B of a CER is the average (sample mean) of the "residuals," namely the differences between the cost estimates and their respective actual costs, corresponding to

all points in the supporting data base. In the OLS context, the bias always turns out to be zero, viz.

$$\begin{aligned} B &= \frac{1}{n}\sum_{k=1}^{n}(a + bx_k - y_k) = \frac{1}{n}\sum_{k=1}^{n}a + \frac{1}{n}b\sum_{k=1}^{n}x_k - \frac{1}{n}\sum_{k=1}^{n}y_k \\ &= \frac{1}{n}na + b\left(\frac{1}{n}\sum_{k=1}^{n}x_k\right) - \frac{1}{n}\sum_{k=1}^{n}y_k = a - \left(\frac{1}{n}\sum_{k=1}^{n}y_k - b\frac{1}{n}\sum_{k=1}^{n}x_k\right) = a - a = 0. \end{aligned}$$

Finally, R^2, often called the coefficient of determination, is the square of the Pearson correlation between the cost estimates and their respective actual costs, corresponding to all points in the supporting data base. R^2 indicates the proportion of variation in the costs that is attributable to the OLS linear relationship between costs and cost drivers. It is usually expressed as a percentage between 0% and 100%. An R^2 of 80%, for example, means that 80% of the variation in the cost values seen in the data base is attributable to variations in the cost-driver values, while the remaining 20% of the variation is attributable to other factors not taken account of in the model, typically additional unidentified cost drivers.

Weighted Least Squares

Weighted least-squares (WLS) regression allows the cost analyst to take into account, not only the historical-cost data themselves, but also the data-collection or estimating context within which the data were gathered or the use to which any resulting CER will be put. Sometimes, the analyst will know that certain data points are less reliably known than others, so he or she can "deweight" the less reliable ones. Sometimes, the analyst will need a CER that estimates cost only within a certain cost-driver range, and then he or she can deweight data points outside that range. Once WLS theory is understood, further application contexts will almost certainly present themselves.

In addition to the actual values of cost driver and cost, each data point is assigned a weight, based on considerations discussed above, so that the set of data consist of triples (x_k,y_k,w_k), where the weight w_k represents the influence that the data point (x_k,y_k) is to have on the CER derived from the data set. In WLS regression, we weight each squared difference $d_k^2 = (y_k - (a+bx_k))^2 = (y_k - a - bx_k)^2$ by its weight w_k. We may

express the principle of weighted least squares as choosing the numerical values of the coefficients $\boldsymbol{a}$ and $\boldsymbol{b}$ by minimizing the weighted sum of squared errors:

$$g(a,b) = \sum_{k=1}^{n} w_k d_k^2 = \sum_{k=1}^{n} w_k (y_k - a - bx_k)^2 \; .$$

What effect on the numerical values of $\boldsymbol{a}$ and $\boldsymbol{b}$ does the weighting procedure have? Well, suppose a particular value w_k is "small," indicating that we do not want the data point (x_k,y_k) to exert a major influence on the CER. Then, regardless of the choice of $\boldsymbol{a}$ and $\boldsymbol{b}$, the term $w_k (y_k - a - bx_k)^2$ is not going to contribute too much to the sum of squared errors. Therefore, the mathematics does not have to move the regression line too close to the data point (x_k,y_k) in order to minimize the sum, because not much will be gained by making an already small summand a little smaller. On the other hand, suppose w_k is "large," indicating that we do want the corresponding data point (x_k,y_k) to exert a major influence on the CER. In this case, the term $w_k (y_k - a - bx_k)^2$ will be a major contributor to the sum of squared errors. In order to make the sum of squared errors as small as possible, $\boldsymbol{a}$ and $\boldsymbol{b}$ will have to be selected to push the resulting CER very close to the point (x_k,y_k).

Normalizing the Weights

Given an initial set of weights $\{ w_1^*, w_2^*, \ldots, w_n^* \}$, we can define a new set of weights $\{w_1, w_2, \ldots, w_n\}$ that is equivalent to the initial set in the sense that the relative weights of all data points are the same as they were, but such that $\sum_{k=1}^{n} w_k = n.$ The new weights are defined, for each $j = 1, 2, \ldots, n$, as $w_j = \dfrac{n w_j^*}{\sum_{k=1}^{n} w_j^*}$. Notice that, for all i and j values, the ratio $\dfrac{w_i}{w_j}$ is the same as the ratio $\dfrac{w_i^*}{w_j^*}$, i.e., the relative values of the new weights with respect to each are the same as the relative values of the original weights with respect to each other. In the sequel, we shall therefore consider all sets $\{w_1,w_2,\ldots,w_n\}$ of weights to be "normalized" in the sense that $\sum_{k=1}^{n} w_k = n.$ Normalization plays a role in simplifying the expressions for the regression coefficients $\boldsymbol{a}$ and $\boldsymbol{b}$, as is shown in the next section.

Derivation of WLS Regression Coefficients

To obtain the mathematical expression for ***a*** and ***b*** in the WLS context, we apply calculus to minimize the weighted sum of squared errors ***g(a,b)*** by first taking the partial derivatives with respect to ***a*** and ***b***:

$$\frac{\partial g}{\partial a} = \sum_{k=1}^{n} 2w_k\left(y_k - a - bx_k\right)(-1) = -2\left(\sum_{k=1}^{n} w_k y_k - a\sum_{k=1}^{n} w_k - b\sum_{k=1}^{n} w_k x_k\right)$$

and

$$\frac{\partial g}{\partial b} = \sum_{k=1}^{n} 2w_k\left(y_k - a - bx_k\right)(-x_k) = -2\left(\sum_{k=1}^{n} w_k x_k y_k - a\sum_{k=1}^{n} w_k x_k - b\sum_{k=1}^{n} w_k x_k^2\right)$$

Setting the two partial derivatives equal to ***0***, we obtain the following two simultaneous equations in the unknowns ***a*** and ***b***:

$$a\sum_{k=1}^{n} w_k + b\sum_{k=1}^{n} w_k x_k = \sum_{k=1}^{n} w_k y_k$$

$$a\sum_{k=1}^{n} w_k x_k + b\sum_{k=1}^{n} w_k x_k^2 = \sum_{k=1}^{n} w_k x_k y_k \ .$$

The solution to these equations is

$$b = \frac{\left(\sum_{k=1}^{n} w_k\right)\left(\sum_{k=1}^{n} w_k x_k y_k\right) - \left(\sum_{k=1}^{n} w_k x_k\right)\left(\sum_{k=1}^{n} w_k y_k\right)}{\left(\sum_{k=1}^{n} w_k\right)\left(\sum_{k=1}^{n} w_k x_k^2\right) - \left(\sum_{k=1}^{n} w_k x_k\right)^2}$$

$$a = \frac{\left(\sum_{k=1}^{n} w_k y_k\right)}{\left(\sum_{k=1}^{n} w_k\right)} - b\frac{\left(\sum_{k=1}^{n} w_k x_k\right)}{\left(\sum_{k=1}^{n} w_k\right)} .$$

Because the weights are normalized, the expressions for ***b*** and ***a*** can be reduced to, respectively,

$$\boldsymbol{b} = \frac{\boldsymbol{n}\left(\sum_{k=1}^{n} w_k x_k y_k\right) - \left(\sum_{k=1}^{n} w_k x_k\right)\left(\sum_{k=1}^{n} w_k y_k\right)}{\boldsymbol{n}\left(\sum_{k=1}^{n} w_k x_k^2\right) - \left(\sum_{k=1}^{n} w_k x_k\right)^2}$$

$$\boldsymbol{a} = \frac{\left(\sum_{k=1}^{n} w_k y_k\right)}{n} - b\frac{\left(\sum_{k=1}^{n} w_k x_k\right)}{n}$$

It is should be noted that when all w_k values are equal (i.e., all equal to 1 assuming normalization), the WLS expressions for a and b reduce to the OLS expressions. In addition, we refer to the expressions

$$\bar{x}_w = \frac{\left(\sum_{k=1}^{n} w_k x_k\right)}{n} \quad \text{and} \quad \bar{y}_w = \frac{\left(\sum_{k=1}^{n} w_k x_k\right)}{n}$$

as the "weighted means" of the x and y values, respectively. Note that the expression for a guarantees that the point $(\bar{x}_w, \bar{y}_w)$ falls exactly on the WLS regression line. Again, when each $w_k = 1$ or, more specifically, when all w_k values are equal, the expressions for the weighted means reduce to the expressions for the ordinary means (i.e., the averages) of x and y.

WLS CER Quality Metrics

The same three quality metrics used for OLS allow the cost analyst to assess the applicability of the WLS CER to estimating problems involving the kinds of subsystems and/or components of which the supporting data base is comprised and the validity of estimates made using it. These three quality metrics are again the following: (1) standard error of the estimate SEE_w; (2) bias B_w; and (3) R_w^2. However, as one would expect, the formulas for them are slightly different in the WLS situation.

Because there is nothing in the WLS setup that plays the OLS role of σ, we consider the standard error of the estimate SEE_w to measure the closeness of the estimated costs $a + bx_k$ to the actual costs y_k in the data base. Its expression is

$$SEE_w = \sqrt{\frac{\sum_{k=1}^{n} w_k (y_k - a - bx_k)^2}{\sum_{k=1}^{n} w_k - 2}} = \sqrt{\frac{\sum_{k=1}^{n} w_k y_k^2 - a\sum_{k=1}^{n} w_k y_k - b\sum_{k=1}^{n} w_k x_k y_k}{n-2}}.$$

In the WLS context, SEE_w is expressed in the same units as the costs and cost estimates, usually dollars. Because the coefficients of the WLS CER are calculated by minimizing the numerator under the square-root sign, the smaller SEE_w turns out to be, the "better" the CER is. Because the weights are normalized, the denominator reduces to $n-2$. If all weights are equal, SEE_w reduces to the unbiased form of the OLS SEE.

The bias B_w of a CER is the weighted mean of the "residuals," namely the differences between the cost estimates and their respective actual costs, corresponding to all points in the supporting data base. As noted earlier, in the OLS context, the bias always turns out to be zero, but this is not true in the WLS context.

$$\begin{aligned}
B_w &= \frac{1}{n}\sum_{k=1}^{n}(a + bx_k - y_k) = \frac{1}{n}\sum_{k=1}^{n} a + \frac{1}{n} b\sum_{k=1}^{n} x_k - \frac{1}{n}\sum_{k=1}^{n} y_k \\
&= \frac{1}{n} na + b\left(\frac{1}{n}\sum_{k=1}^{n} x_k\right) - \frac{1}{n}\sum_{k=1}^{n} y_k = a + b\left(\frac{1}{n}\sum_{k=1}^{n} x_k\right) - \frac{1}{n}\sum_{k=1}^{n} y_k \\
&= a - \frac{\left(\sum_{k=1}^{n} y_k\right) - b\left(\sum_{k=1}^{n} x_k\right)}{n} \quad a = \frac{\left(\sum_{k=1}^{n} w_k y_k\right)}{n} - b\frac{\left(\sum_{k=1}^{n} w_k x_k\right)}{n} - \frac{\left(\sum_{k=1}^{n} y_k\right) - b\left(\sum_{k=1}^{n} x_k\right)}{n} \\
&= \frac{\left(\sum_{k=1}^{n} w_k y_k\right)}{n} - \frac{\left(\sum_{k=1}^{n} y_k\right)}{n} - b\frac{\left(\sum_{k=1}^{n} w_k x_k\right)}{n} + \frac{\left(\sum_{k=1}^{n} x_k\right)}{n} \\
&= \frac{\left(\sum_{k=1}^{n} (w_k - 1) y_k\right)}{n} - b\frac{\left(\sum_{k=1}^{n} (w_k - 1) x_k\right)}{n}
\end{aligned}$$

which reduces to 0 when all $w_k = 1$ or, more specifically, are all the same when normalized. However, the bias is, in general, not typically zero in the weighted least-squares situation.

Finally, R^2, just as in the OLS situation, measures the worth of the linear-regression equation as a model of the relationship underlying the data base. To derive the formula for R^2 in the WLS situation, let's start with some reasoning that applies in the

OLS situation. Referring to the data points (x_1,y_1), (x_2,y_2), …, (x_n,y_n), we ask why the y values vary, i.e., why are they not all the same. There are two basic reasons that the y values vary: (1) the x values vary, and y is related to x through the hypothesized linear relationship, and (2) any other reason you can think of that does not involve the hypothesized linear relationship, e.g., nonlinearity, random errors in the data, additional cost drivers, that affects y. What R^2 does is to allocate the variation in y between these two sources. In particular R^2, usually expressed as a percentage, indicates the proportion of variation in y that is attributable to the linear relationship between x and y.

If the y values did not vary at all from the WLS regression line, they all would be equal to their weighted mean $\bar{y}_w = \left(\sum_{k=1}^{n} w_k y_k\right) / n$. If, on the other hand, we had no knowledge at all about the relationship between x and y, the best we could do to predict the value y at any given x would be to predict $y = \bar{y}_w$. This is equivalent to using the horizontal line $y = \bar{y}_w$ in place of the regression line $y = a + bx$. The sum of squared errors from the horizontal line $y = \bar{y}_w$ is called the "total variation" of y and is denoted $TV = \sum_{k=1}^{n} w_k (y_k - \bar{y}_w)^2$.

Suppose now that the only variation in y were due to the influence of the regression line $y = a + bx$. Then every y_k would be equal to its corresponding $a+bx_k$. The resulting total variation would then be

$$\sum_{k=1}^{n} w_k (y_k - \bar{y}_w)^2 = \sum_{k=1}^{n} w_k (a + bx_k - \bar{y}_w)^2$$

since each y_k and $a+bx_k$ would be one and the same. It would follow that the quantity $VR = \sum_{k=1}^{n} w_k (a + bx_k - \bar{y}_w)^2$, called the "variance due to regression" is the variation in y that can be attributed to the impact of the regression relationship.

We then compare TV and VR with the weighted sum of squared (SS) errors, where $\mathrm{SS} = \sum_{k=1}^{n} w_k (y_k - a - bx_k)^2$. It can be proved by elementary, though tedious,

calculations that ***TV = SS+VR***. These calculations are reproduced in the Appendix. Simple algebra then ensures that $\frac{SS}{TV}+\frac{VR}{TV}=1.$ From this equation, it is evident that VR/TV is the proportion of the total variation in y that can be attributed to the impact of the linear-regression relationship. The proportion of variation in y due to all other effects is equal to SS/TV. The WLS coefficient of determination is then

$$R_w^2 = \frac{VR}{TV} = \frac{\sum_{k=1}^{n} w_k (a+bx_k-\bar{y}_w)^2}{\sum_{k=1}^{n} w_k (y_k-\bar{y}_w)^2} = \frac{\sum_{k=1}^{n} w_k (a+bx_k-a-b\bar{x}_w)^2}{\sum_{k=1}^{n} w_k (y_k^2-2y_k\bar{y}_w+\bar{y}_w^2)}$$

$$= \frac{b^2\sum_{k=1}^{n} w_k (x_k-\bar{x}_w)^2}{\sum_{k=1}^{n} w_k y_k^2 - 2\bar{y}_w\sum_{k=1}^{n} w_k y_k + n\bar{y}_w^2} = \frac{b^2\left(\sum_{k=1}^{n} w_k x_k^2 - 2\bar{x}_w\sum_{k=1}^{n} w_k x_k + n\bar{x}_w^2\right)}{\sum_{k=1}^{n} w_k y_k^2 - 2\bar{y}_w\sum_{k=1}^{n} w_k y_k + n\bar{y}_w^2}$$

$$= \frac{b^2\left(\sum_{k=1}^{n} w_k x_k^2 - \left(\sum_{k=1}^{n} w_k x_k\right)^2 / n\right)}{\sum_{k=1}^{n} w_k y_k^2 - \left(\sum_{k=1}^{n} w_k y_k\right)^2 / n}$$

$$= \frac{\left\{n\left(\sum_{k=1}^{n} w_k x_k y_k\right) - \left(\sum_{k=1}^{n} w_k x_k\right)\left(\sum_{k=1}^{n} w_k y_k\right)\right\}^2}{\left\{n\left(\sum_{k=1}^{n} w_k x_k^2\right) - \left(\sum_{k=1}^{n} w_k x_k\right)^2\right\}^2} \times \frac{\sum_{k=1}^{n} w_k x_k^2 - \left(\sum_{k=1}^{n} w_k x_k\right)^2 / n}{\sum_{k=1}^{n} w_k y_k^2 - \left(\sum_{k=1}^{n} w_k y_k\right)^2 / n}$$

$$R_w^2 = \frac{\left\{n\left(\sum_{k=1}^{n} w_k x_k y_k\right) - \left(\sum_{k=1}^{n} w_k x_k\right)\left(\sum_{k=1}^{n} w_k y_k\right)\right\}^2}{\left\{n\left(\sum_{k=1}^{n} w_k x_k^2\right) - \left(\sum_{k=1}^{n} w_k x_k\right)^2\right\}\left\{n\left(\sum_{k=1}^{n} w_k y_k^2\right) - \left(\sum_{k=1}^{n} w_k y_k\right)^2\right\}}$$

Adaptive CERs via Quadratic-Distance Weighting

An "adaptive" CER is an extension of the concept of analogy estimating to the CER context. The standard way doing analogy estimating is by finding one historical program that has several characteristics in common with the subsystems or components of a program that is being estimated, for example, the program's objective, hardware or software design proposed to carry it out, materials of which any hardware is constructed, use of similar legacy components, and Government or contractor approach to program development or production. The idea behind an adaptive CER is to build a data base consisting of as many programs as we can find that have subsystems or components of the same basic kind as in the program being estimated. Normally, we would use all the points of this data base to derive a CER that expresses the subsystem or component cost in terms of an appropriate cost-driver.

However, in any particular estimating context, we are interested only in one particular value of the cost driver or, at most, a relatively short interval of such values. We know from classical OLS theory (see below) that, if the value at which we are interested in estimating is relatively far away from the cost-driver values in the data base, the accuracy of our estimate is substantially reduced. Adaptive CERs look at the flip side of this situation: If a cost-driver value of a data point is relatively far away from the value at which we want to do our estimate, maybe we don't want to use that data point to calculate our CER or, at least, maybe we don't want to consider it of equal weight with data points whose cost-driver values are closer to where we want to estimate.

The mechanics of calculating adaptive CERs is therefore based on measurements of the distance between cost-driver values in the data base and the cost-driver value at which we want to conduct our estimate. Data points are treated differently, according to their distance from the estimating point. To carry out the process, we assign each point in the data base a "weight" that indicates how important that data point is to our estimating problem. Then we apply "weighted least-squares" (WLS) regression to derive the CER.

For purposes of illustration in this paper, we shall consider quadratic-distance weighting. This weighting method calls for weighting points according to the squared distance of its cost-driver value along the x-axis from a cost-driver value of interest. If $\boldsymbol{x_0}$ is the cost-driver value of interest and $\boldsymbol{x_k}$ is the cost-driver value of the $\boldsymbol{k^{th}}$ data point, then

$QD_k = (x_0 - x_k)^2$ is the squared distance between the two cost-driver values. Because the greater that distance is, the less we want its weight to be, we define the weight of the data point (x_k, y_k) to be the reciprocal of QD_k, namely $w_k = (x_0 - x_k)^{-2}$.

Why choose quadratic-distance weighting from among the infinite number of ways to define the weighting in terms of a cost driver's distance from x_0? We prefer the squared (quadratic) distance, because OLS calculations use the squares of residuals for best fit – this process forces the CER to pass through the point $(\bar{x}, \bar{y})$, where $\bar{x}$ is the mean of the cost-driver values and $\bar{y}$ is the mean of the cost values in the data base. In the WLS case, the regression line based on minimizing the squares of residuals passes through the point $(\bar{x}_w, \bar{y}_w)$, where $\bar{x}_w = \left(\sum_{k=1}^{k} w_k x_k\right) \div \left(\sum_{k=1}^{k} w_k\right)$ is the weighted mean of the cost-driver values and $\bar{y}_w = \left(\sum_{k=1}^{k} w_k y_k\right) \div \left(\sum_{k=1}^{k} w_k\right)$ is the weighted mean of the cost values. However, other weighting schemes can be used if there is a compelling reason to do so.

Starting with the historical-cost data in Table 22, suppose we want to estimate the cost of a similar subsystem or component of interest whose cost-driver value is 800. We then weight each of the data points according to the quadratic distance of its cost-driver value from 800. The results are listed in Table 23. Note that the normalized weights sum to 19, which is the number of data points.

Program	Cost-Driver Value x	Unit Cost y	Initial Weight w	Normalized Weight w
A	156.12	51,367.22	0.00000241	0.003881827
B	179.40	5,885.00	0.00000260	0.004178521
C	180.30	7,060.00	0.00000260	0.004190667
D	217.50	139,483.12	0.00000295	0.004743012
E	419.14	3,386.00	0.00000689	0.011094695
F	437.09	6,738.00	0.00000759	0.012219353
G	440.93	6,812.00	0.00000776	0.012482106
H	494.45	3,291.34	0.00001071	0.017237787
I	789.90	5,723.14	0.00980296	15.77623429
J	826.10	10,992.00	0.00146798	2.362463352
K	864.30	11,590.00	0.00024187	0.389245992
L	869.30	15,973.00	0.00020823	0.335104011
M	976.50	7,970.67	0.00003210	0.05166027
N	1,355.80	9,524.10	0.00000324	0.005209656
O	1,360.90	35,927.22	0.00000318	0.005115348
P	1,463.21	11,238.73	0.00000227	0.003658845
Q	2,332.10	92,059.97	0.00000043	0.000685602
R	3,017.73	74,649.00	0.00000020	0.000327212
S	3,253.00	42,915.23	0.00000017	0.000267455
Sums	19,633.77	542,585.74	0.01180613	19.00000000

Table 23. Historical-Cost Data Weighted According to their Quadratic Distances from 800

The next step is to calculate the adaptive CER, i.e., the CER adapted to estimating at a cost-driver value of 800. We apply WLS methods to derive this CER, i.e., using the formulas for ***a*** and ***b*** derived earlier. The required preliminary computations appear in Table 24.

Cost-Driver Value of Interest =			800						
Program	Cost-Driver Value x	Unit Cost y	Normalized Weight w	wx	wy	wx^2	wy^2	wxy	WLS EST y
A	156.12	51,367.22	0.00388183	0.61	199.40	94.61	10,242,556.04	31,130.12	-5,734.14
B	179.40	5,885.00	0.00417852	0.75	24.59	134.48	144,715.65	4,411.55	-5,280.91
C	180.30	7,060.00	0.00419067	0.76	29.59	136.23	208,877.91	5,334.37	-5,263.39
D	217.50	139,483.12	0.00474301	1.03	661.57	224.37	92,277,865.87	143,891.50	-4,539.16
E	419.14	3,386.00	0.01109470	4.65	37.57	1,949.10	127,200.63	15,745.68	-613.53
F	437.09	6,738.00	0.01221935	5.34	82.33	2,334.48	554,766.51	35,987.37	-264.07
G	440.93	6,812.00	0.01248211	5.50	85.03	2,426.76	579,211.44	37,491.44	-189.31
H	494.45	3,291.34	0.01723779	8.52	56.74	4,214.31	186,735.55	28,052.83	852.64
I	789.90	5,723.14	15.77623429	12,461.65	90,289.60	9,843,455.33	516,740,007.13	71,319,753.08	6,604.62
J	826.10	10,992.00	2.36246335	1,951.63	25,968.20	1,612,242.35	285,442,423.23	21,452,327.68	7,309.38
K	864.30	11,590.00	0.38924599	336.43	4,511.36	290,772.40	52,286,674.49	3,899,169.35	8,053.07
L	869.30	15,973.00	0.33510401	291.31	5,352.62	253,232.23	85,497,341.14	4,653,029.40	8,150.42
M	976.50	7,970.67	0.05166027	50.45	411.77	49,260.77	3,282,058.62	402,090.44	10,237.44
N	1,355.80	9,524.10	0.00520966	7.06	49.62	9,576.36	472,559.94	67,271.11	17,621.85
O	1,360.90	35,927.22	0.00511535	6.96	183.78	9,473.87	6,602,713.33	250,106.54	17,721.14
P	1,463.21	11,238.73	0.00365884	5.35	41.12	7,833.53	462,145.19	60,168.32	19,712.96
Q	2,332.10	92,059.97	0.00068560	1.60	63.12	3,728.78	5,810,500.30	147,193.92	36,628.96
R	3,017.73	74,649.00	0.00032721	0.99	24.43	2,979.82	1,823,378.13	73,711.14	49,977.16
S	3,253.00	42,915.23	0.00026746	0.87	11.48	2,830.21	492,576.73	37,337.61	54,557.52
Sums	19,633.77	542,585.74	19.00000000	15,141.45	128,083.89	12,096,899.99	1,063,234,307.81	102,664,203.45	215,542.66
			Num b =	11,243,876.6334		Std Error =	3,147.8208		
			Den b =	577,541.5425		Num R^2 =	126,424,761,747,155.0000		
			b =	19.4685		Den R^2 =	2,192,330,157,360,000.0000		
			Wtd Mean x =	796.9185		R^2 =	5.7667%		
			Wtd Mean y =	6,741.2572					
			a =	-8,773.5633					

Table 24. WLS Computations Leading to Adaptive CER at a Cost-Driver Value of 800

Figure 2 compares the full-data-set CER with the CER adapted, via quadratic-distance weighting, to a cost-driver value of ***800***. It should be noticed that the standard error of the full-data-set CER is ***34,336.83***, while the standard error of the adaptive CER with points far from ***800*** deweighted considerably is only ***3,147.82***, a decrease in magnitude of over 90 percent.

Note also that the adaptive CER ***y = -8,773.56 + 19.4685x*** appears to estimate more accurately around ***x = 800***, while essentially ignoring data points whose ***x*** values are far removed from ***800***. This view is supported by the relative values of the standard errors of both CERs.

For additional illustration, we compare in Figure 3 the full-data-set CER with the CER adapted, via quadratic-distance weighting, to a cost-driver value of ***300***. It is still true, of course, that the standard error of the full-data-set CER is ***34,336.83***, while the standard error of the adaptive CER with points far from 300 deweighted considerably and those near 300 more heavily weighted is now ***55,556.56***. This large standard error undoubtedly occurs, because the actual data points vary quite a bit near the ***300*** cost-

driver value. In Figure 4, we compare the full-data-set CER with the CER adapted, via quadratic-distance weighting, to a cost-driver value of ***3,000***. While the standard error of the full-data-set CER remains at ***34,336.83***, the standard error of the adaptive CER with points far from 3,000 deweighted is now ***2,838.37***.

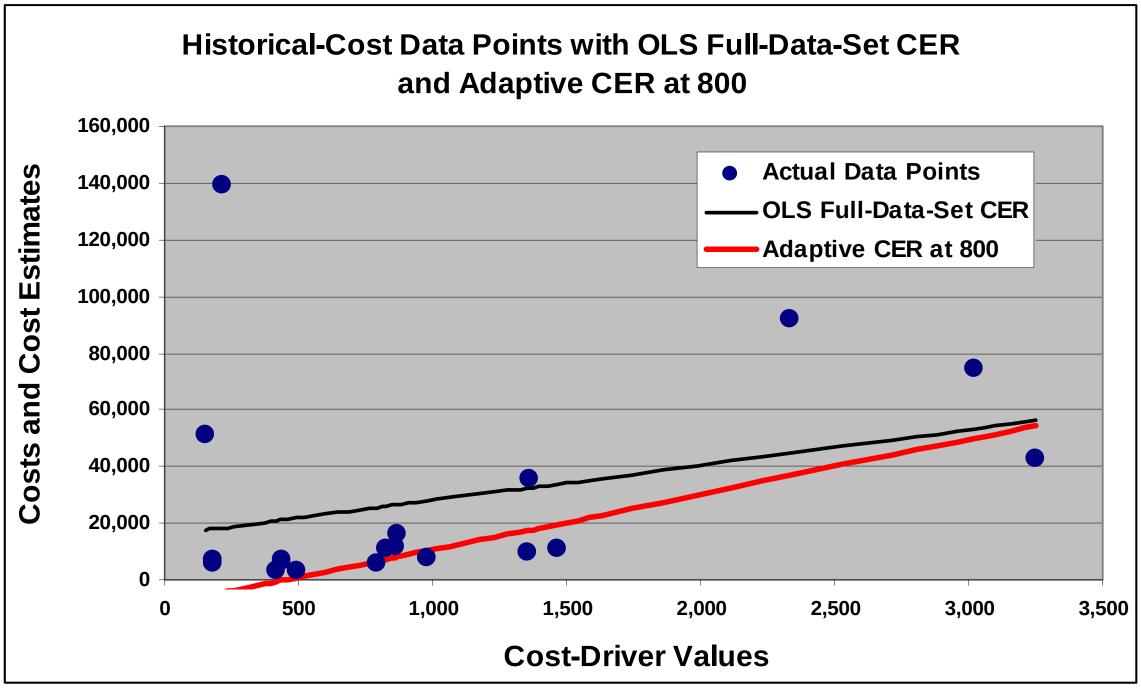

Figure 2. OLS Full-Data-Set CER Compared with Adaptive CER at a Cost-Driver Value of 800

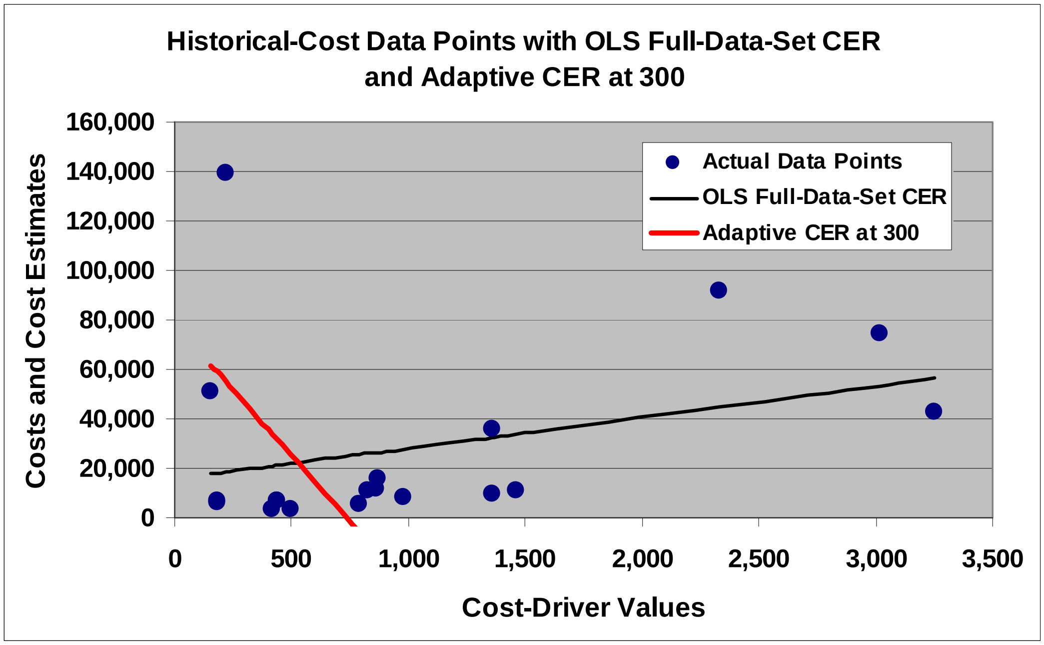

Figure 3. OLS Full-Data-Set CER Compared with Adaptive CER at a Cost-Driver Value of 300

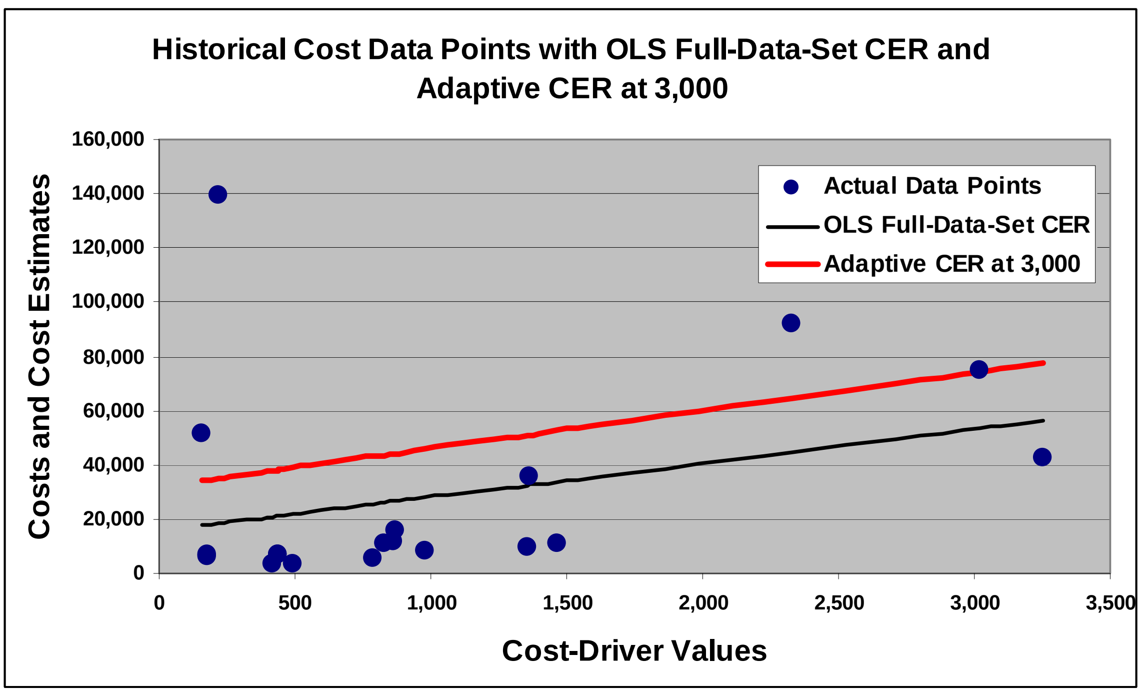

Figure 4. OLS Full-Data-Set CER Compared with Adaptive CER at a Cost-Driver Value of 3,000

The "Universal Adaptive CER"

The "universal adaptive CER" is formed by combining* the various individual adaptive CERs, of the sort derived above, over the range of cost drivers into one CER that applies over the entire range. This "universal adaptive CER" is, as P. Foussier (Reference 3, Chart 5) presciently noted, "highly nonlinear." For the data set we have been working with, we can consider the cost-driver range to go from 50 to 3,500, and we calculate a quadratic-distance-weighted CER and an estimated cost at each increment of 50 for each of those cost-driver values. Then we string all these estimates together and interpolate between successive ones to form the universal adaptive CER.

To complete the picture of estimating at each point along the cost-driver axis, we record and graph the standard error at each point as well. Table 25 contains the estimates and standard errors at 50 units apart along the cost-driver axis. The numbers in Table 25 form the basis for the graphs of the universal adaptive CER and the corresponding standard errors in Figure 5. For comparison purposes, the standard error of the OLS CER is a constant ***34,336.83*** across the database. Notice how the standard error of the universal adaptive CER varies with the distance of the cost-driver value (x axis) from the nearest point in the data base. The numbers in red (between the 50-unit points) in Table 25 identify the actual data points underlying the analysis.

The idea of combining estimates at various points of the cost-driver range into one all-inclusive CER was suggested to us by Paul Wetzel of OpsConsulting LLC.

Driver	EST Cost	Std Error
50.00	42,739.31	46,098.71
100.00	40,817.29	41,490.92
150.00	49,546.82	15,013.91
156.12	50,880.53	20,862.57
179.40	55,953.88	43,110.41
180.30	56,150.02	43,970.50
200.00	60,443.18	62,797.07
217.50	69,749.17	63,712.78
250.00	87,031.73	65,413.39
300.00	46,425.71	57,676.55
350.00	22,733.56	36,873.63
400.00	7,006.95	11,986.04
419.14	6,760.42	9,109.80
437.09	6,529.22	6,412.39
440.93	6,479.76	5,835.34
450.00	6,362.94	4,472.36
494.45	3,589.46	3,084.58
500.00	3,243.16	2,911.31
550.00	6,829.12	17,776.83
600.00	9,959.40	22,010.11
650.00	11,310.17	21,033.96
700.00	10,929.01	16,492.92
750.00	8,652.67	9,456.12
789.90	7,175.24	4,565.75
800.00	6,801.25	3,327.84
826.10	9,756.59	3,386.63
850.00	12,462.82	3,440.47
864.30	12,666.50	4,059.71
869.30	12,737.72	4,276.23
900.00	13,174.99	5,605.64
950.00	9,208.15	5,651.88
976.50	8,832.68	5,342.38
1,000.00	8,499.71	5,067.91
1,050.00	11,462.16	11,841.54
1,100.00	14,296.49	15,323.02
1,150.00	16,537.15	16,912.27
1,200.00	18,230.99	17,020.52
1,250.00	19,495.31	16,029.95
1,300.00	20,310.23	14,631.94
1,350.00	14,974.31	11,522.07
1,355.80	15,774.27	11,821.74
1,360.90	16,477.67	12,085.24
1,400.00	21,870.45	14,105.41
1,450.00	11,840.86	4,214.92
1,463.21	12,101.01	5,274.84
1,500.00	12,825.54	8,226.72
1,550.00	16,621.72	13,974.93
1,600.00	20,492.26	17,569.25
1,650.00	24,526.56	20,350.34
1,700.00	28,831.03	22,668.31
1,750.00	33,415.50	24,632.61
1,800.00	38,247.16	26,275.33
1,850.00	43,285.50	27,589.48
1,900.00	48,497.85	28,534.71
1,950.00	53,862.57	29,032.00
2,000.00	59,364.10	28,954.26
2,050.00	64,981.01	28,118.23
2,100.00	70,666.52	26,286.86
2,150.00	76,319.27	23,197.58
2,200.00	81,744.09	18,634.07
2,250.00	86,609.89	12,543.91
2,300.00	90,430.47	5,163.31
2,332.10	91,836.14	3,730.10
2,350.00	92,619.98	2,930.89
2,400.00	92,676.25	10,907.76
2,450.00	90,463.37	17,895.26
2,500.00	86,410.39	23,227.16
2,550.00	81,412.53	26,603.62
2,600.00	76,466.46	28,091.64
2,650.00	72,322.92	27,995.50
2,700.00	69,366.76	26,697.11
2,750.00	66,431.86	24,540.98
2,800.00	67,242.40	21,772.29
2,850.00	67,904.22	18,495.58
2,900.00	69,545.45	14,613.82
2,950.00	71,913.26	9,720.21
3,000.00	74,219.40	3,000.69
3,017.73	74,164.83	4,164.89
3,050.00	74,065.53	6,283.82
3,100.00	67,141.02	15,848.64
3,150.00	54,415.99	17,689.83
3,200.00	45,424.15	9,943.35
3,250.00	42,927.10	501.90
3,253.00	42,978.65	868.74
3,300.00	43,786.36	6,615.99
3,350.00	45,762.39	11,482.72
3,400.00	47,971.96	14,864.14
3,450.00	50,126.95	17,319.87
3,500.00	52,149.51	19,185.52

Table 25. Universal Adaptive-CER-Based Estimates and Standard Errors at 50-Unit Increments Along the Cost-Driver Axis

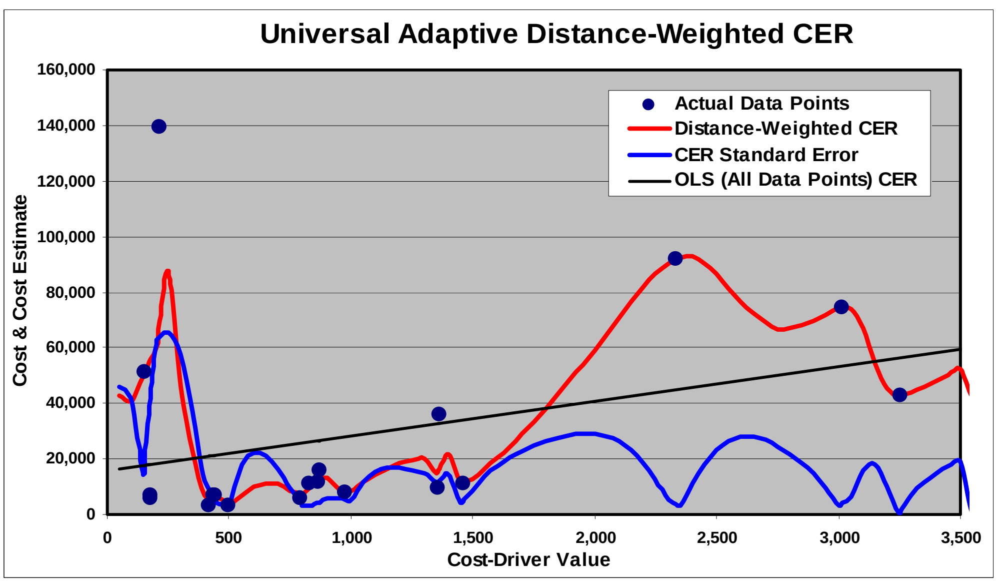

Figure 5. Universal Adaptive-CER-Based Estimates and Standard Errors Graphed at 50-Unit Increments along the Cost-Driver Axis Prediction Bounds

Estimating the cost of developing or producing a new subsystem or component is essentially trying to predict the future, which means that any such estimate contains uncertainty. A portion of this uncertainty is described by the "standard error of the estimate" of a cost-estimating relationship (CER), which is basically the standard deviation of errors made (the "residuals") in using that CER to estimate the (known) costs of the subsystems or components comprising the supporting historical data base. The standard error of the estimate depends primarily on the extent to which those (known) costs fit the CER that purports to model them. However, additional uncertainty arises from the location of the particular cost-driver value (***x***) within or without the range of cost-driver values for programs comprising the historical cost data base. For example, if ***x*** were located near the center of the range of its historical values, the CER would provide a more precise measure of the element's cost than if ***x*** were located far from the center of the range. The total uncertainty in the estimate can then be expressed in terms of prediction bounds that involve both sources of uncertainty.

The first kind of uncertainty, represented by only one number characteristic of the CER, is fairly easy to measure for any CER shape or error model. The second kind, which involves both the CER itself and the value of the cost-driving parameter, however, is more complicated, and the way to calculate it is completely understood only in the case of classical OLS linear regression. As a result, an explicit formula exists for "prediction intervals" that bound cost estimates based on CERs that have been derived by applying OLS to historical cost data. In fact, the formula for the $(1-\alpha)^{th}$ percent upper and lower prediction bounds on the true cost y, based on the estimate ***ESTy*** from the CER is the following:

$$ESTy \pm t_{\alpha/2,n-2} * SEE\sqrt{1+\frac{1}{n}+\frac{(x-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}$$

where $t_{\alpha/2,n-2}$ is the $(1-\alpha)^{th}$ percentage point of the t distribution, $\bar{x}$ is the mean of the cost-driver values in the data base, x is the cost-driver value at which the estimate is being

made, and ***SEE*** is the standard error of the estimate. Table 26 displays the sequence of 80% upper and lower prediction bounds for the OLS CER based on our data set. Figure 6 graphs the prediction bounds, along with the actual data points and the OLS CER.

Program	Cost-Driver Value x	Unit Cost y	80% Upper Bound	OLS EST y	80% Lower Bound
A	156.12	51,367.22	65,673.53	17,596.30	-30,480.93
B	179.40	5,885.00	65,907.23	17,887.18	-30,132.88
C	180.30	7,060.00	65,916.29	17,898.42	-30,119.45
D	217.50	139,483.12	66,292.88	18,363.23	-29,566.43
E	419.14	3,386.00	68,400.42	20,882.67	-26,635.08
F	437.09	6,738.00	68,593.51	21,106.95	-26,379.62
G	440.93	6,812.00	68,634.94	21,154.93	-26,325.09
H	494.45	3,291.34	69,216.65	21,823.65	-25,569.35
I	789.90	5,723.14	72,574.56	25,515.22	-21,544.12
J	826.10	10,992.00	73,003.23	25,967.53	-21,068.17
K	864.30	11,590.00	73,459.69	26,444.83	-20,570.03
L	869.30	15,973.00	73,519.75	26,507.30	-20,505.14
M	976.50	7,970.67	74,824.83	27,846.74	-19,131.35
N	1,355.80	9,524.10	79,710.04	32,586.00	-14,538.05
O	1,360.90	35,927.22	79,778.56	32,649.72	-14,479.12
P	1,463.21	11,238.73	81,168.85	33,928.06	-13,312.74
Q	2,332.10	92,059.97	94,145.23	44,784.62	-4,576.00
R	3,017.73	74,649.00	105,728.61	53,351.39	974.17
S	3,253.00	42,915.23	109,940.12	56,291.03	2,641.94

Table 26. Eighty Percent Upper and Lower OLS Prediction Bounds

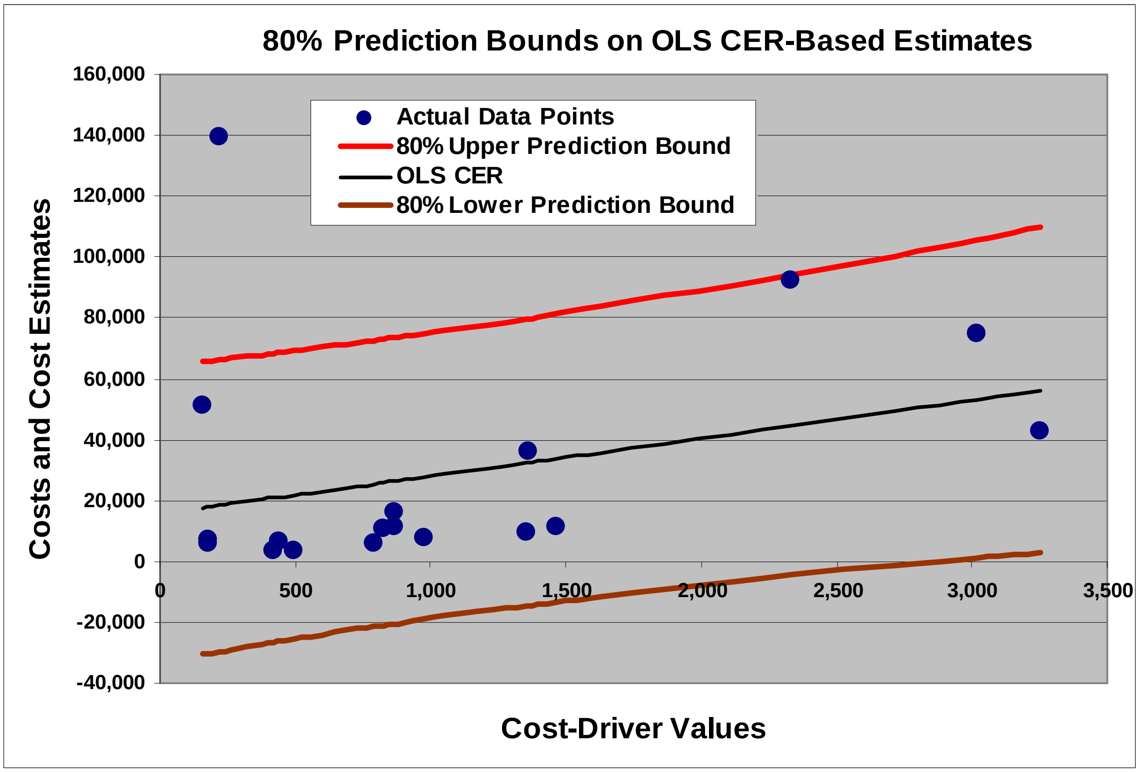

Figure 6. Eighty Percent OLS Prediction Bounds with Actual Data Points and OLS CER

When the weights are normalized, the expressions for the $(1\text{-}\alpha)^{th}$ percent upper and lower prediction bounds on the true cost y at the cost-driver value x_p, based on estimates ***ESTy*** from WLS-based adaptive CERs are the following:

$$ESTy \pm \ t_{\alpha/2,n-2} * SEE_w \sqrt{\frac{1}{w_p} + \frac{1}{n} + \frac{n(x_p - \bar{x})^2}{n\left(\sum_{k=1}^{n} w_k x_k^2\right) - \left(\sum_{k=1}^{k} w_k x_k\right)^2}}$$

One way to obtain a usable value, if needed, for w_p when x_p is not in the data base from which the adaptive CERs are derived is to interpolate between the weights of the nearest data-base points. That is what is effectively done in the graphs based on Tables 6, 7, and 8 below.

In Table 26, 28, and 29, we compile the 80% upper and lower prediction bounds on adaptive CERs at the cost-driver values, respectively, of 800, 300, and 3,000. Figures 7, 8, and 9 display the graphs of these respective prediction bounds. Notice how the prediction bounds narrow in the region very near the cost-driver value of interest.

Program	Cost-Driver Value x	Unit Cost y	80% Upper Bound	WLS EST y	80% Lower Bound
A	156.12	51,367.22	67,335.731428	-5,734.14	-78,804.008697
B	179.40	5,885.00	65,146.948025	-5,280.91	-75,708.771200
C	180.30	7,060.00	65,062.330513	-5,263.39	-75,589.110360
D	217.50	139,483.12	61,564.835765	-4,539.16	-70,643.158038
E	419.14	3,386.00	42,608.518817	-613.53	-43,835.578046
F	437.09	6,738.00	40,921.251654	-264.07	-41,449.391167
G	440.93	6,812.00	40,560.306422	-189.31	-40,938.927733
H	494.45	3,291.34	35,529.986321	852.64	-33,824.697703
I	789.90	5,723.14	8,126.533982	6,604.62	5,082.700610
J	826.10	10,992.00	10,459.318778	7,309.38	4,159.436356
K	864.30	11,590.00	15,439.587849	8,053.07	666.561891
L	869.30	15,973.00	16,099.371097	8,150.42	201.463800
M	976.50	7,970.67	30,313.734118	10,237.44	-9,838.849438
N	1,355.80	9,524.10	80,730.945765	17,621.85	-45,487.245014
O	1,360.90	35,927.22	81,409.009710	17,721.14	-45,966.730098
P	1,463.21	11,238.73	95,011.748000	19,712.96	-55,585.820690
Q	2,332.10	92,059.97	210,542.762967	36,628.96	-137,284.838305
R	3,017.73	74,649.00	301,708.981386	49,977.16	-201,754.659776
S	3,253.00	42,915.23	332,992.265384	54,557.52	-223,877.228359
Sums	19,633.77	542,585.74		215,542.66	

Table 27. Eighty Percent Upper and Lower Prediction Bounds for Adaptive-CER-Based Estimates at Cost-Driver Value 800

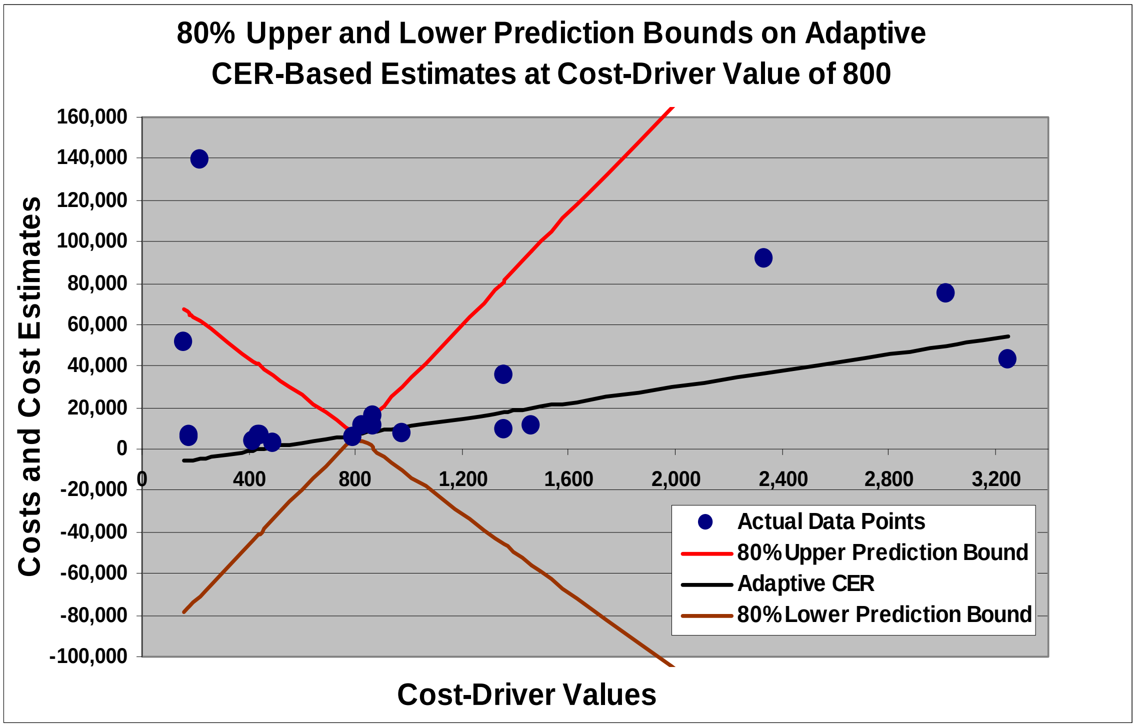

Figure 7. Eighty Percent Prediction Bounds for Adaptive-CER-Based Estimates at Cost-Driver Value 800 with Actual Data Points and Adaptive CER

What is characteristic about the prediction bounds whose graphs appear in Figures 7, 9, and 11 is their excessive widening as the cost-driver value moves away from its base value (800 in Figure 7, 300 in Figure 9, and 3,000 in Figure 11. The point to remember about adaptive CERs is that it is our intention to apply them <u>only</u> in the vicinity of the base cost-driver value, where the prediction bounds are at their narrowest. Therefore, their width in other estimating regions is essentially irrelevant. By the way, the upper and lower prediction bounds do not touch, as Figures 8, 10, and 12 show. In addition, because these are prediction bounds on cost estimates, which as a practical matter cannot be negative, the region of applicability is further constrained beyond cost-driver values at which the lower prediction bounds go negative.

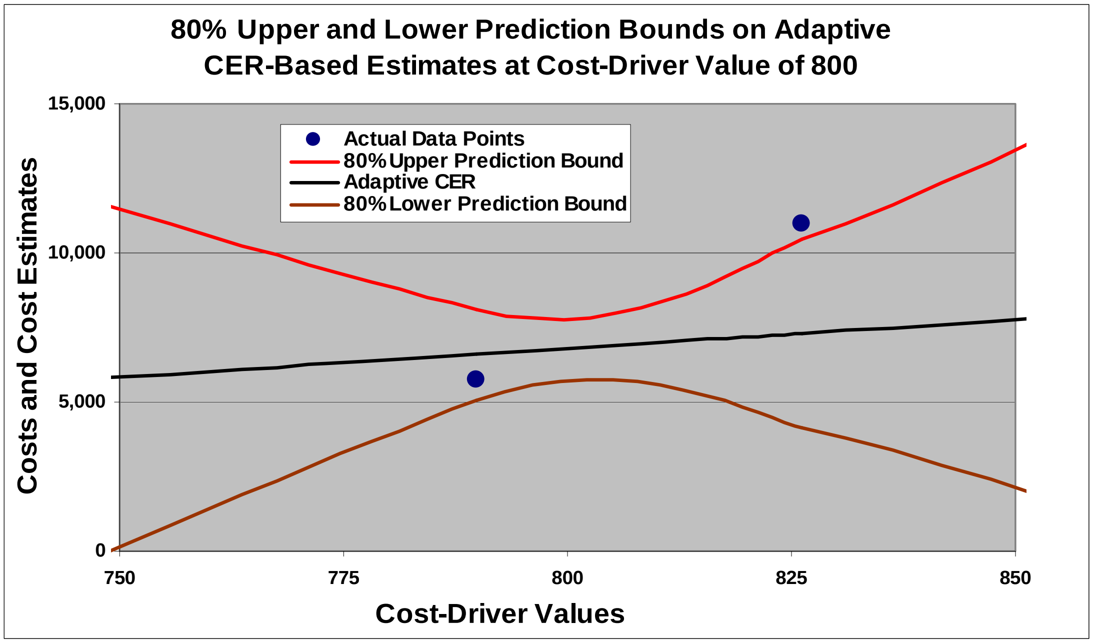

Figure 8. Gap between Upper and Lower Prediction Bounds in the Vicinity of the Cost-Driver Value 800

Program	Cost-Driver Value x	Unit Cost y	80% Upper Bound	WLS EST y	80% Lower Bound
A	156.12	51,367.22	65,389.279544	61,698.97	58,008.663971
B	179.40	5,885.00	62,372.227016	59,227.74	56,083.244080
C	180.30	7,060.00	62,255.776784	59,132.20	56,008.619347
D	217.50	139,483.12	57,462.441876	55,183.32	52,904.189048
E	419.14	3,386.00	36,867.788626	33,778.67	30,689.557986
F	437.09	6,738.00	35,381.736102	31,873.23	28,364.726492
G	440.93	6,812.00	35,064.501531	31,465.60	27,866.707881
H	494.45	3,291.34	30,658.711130	25,784.31	20,909.907048
I	789.90	5,723.14	6,491.040727	-5,578.52	-17,648.087346
J	826.10	10,992.00	3,534.857637	-9,421.25	-22,377.363947
K	864.30	11,590.00	415.759782	-13,476.29	-27,368.336816
L	869.30	15,973.00	7.527753	-14,007.05	-28,021.632368
M	976.50	7,970.67	-8,743.802865	-25,386.63	-42,029.453100
N	1,355.80	9,524.10	-39,698.603983	-65,650.37	-91,602.134324
O	1,360.90	35,927.22	-40,114.762116	-66,191.75	-92,268.734323
P	1,463.21	11,238.73	-48,463.042557	-77,052.24	-105,641.431258
Q	2,332.10	92,059.97	-119,355.526647	-169,287.31	-219,219.087245
R	3,017.73	74,649.00	-175,292.373781	-242,068.82	-308,845.271266
S	3,253.00	42,915.23	-194,486.501042	-267,043.38	-339,600.262830
Sums	19,633.77	542,585.74		-597,019.57	

Table 28. Zero Percent Upper and Lower Prediction Bounds for Adaptive-CER-Based Estimates at Cost-Driver Value 300

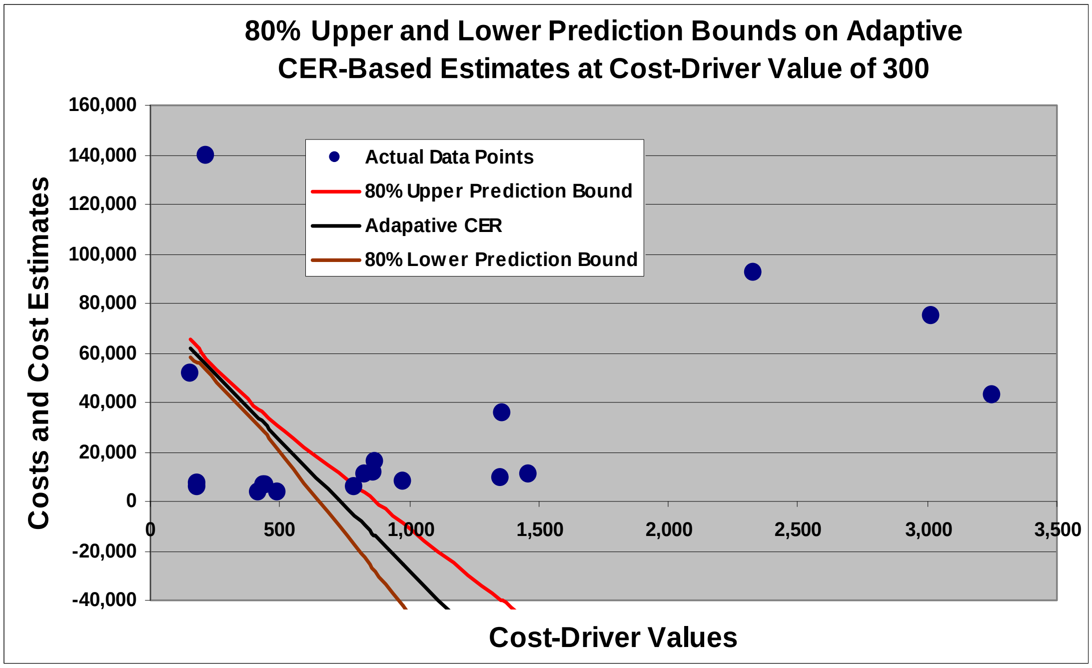

Figure 9. Eighty Percent Prediction Bounds for Adaptive-CER-Based Estimates at Cost-Driver Value 300 with Actual Data Points and Adaptive CER

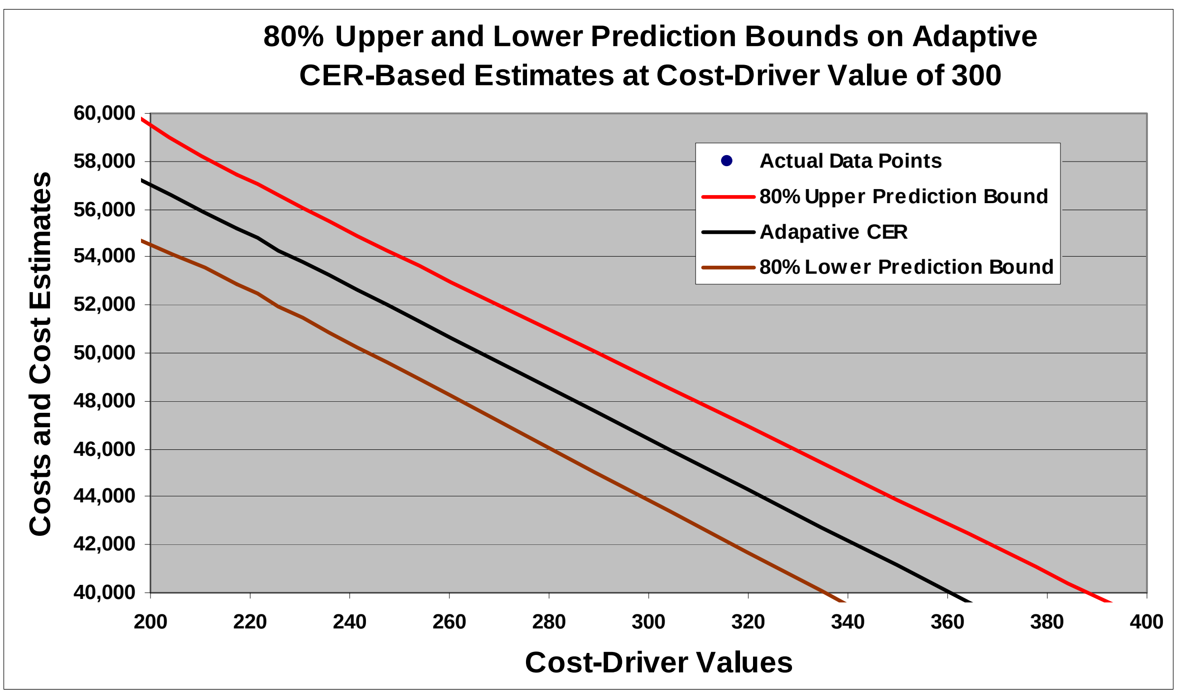

Figure 10. Gap between Upper and Lower Prediction Bounds in the Vicinity of the Cost-Driver Value 300

Program	Cost-Driver Value x	Unit Cost y	80% Upper Bound	WLS EST y	80% Lower Bound
A	156.12	51,367.22	202,434.005312	34,104.71	-134,224.591913
B	179.40	5,885.00	201,384.901034	34,433.09	-132,518.729992
C	180.30	7,060.00	201,344.342887	34,445.78	-132,452.781730
D	217.50	139,483.12	199,667.940092	34,970.51	-129,726.920845
E	419.14	3,386.00	190,581.137616	37,814.77	-114,951.604146
F	437.09	6,738.00	189,772.232090	38,067.96	-113,636.306880
G	440.93	6,812.00	189,599.184936	38,122.13	-113,354.928569
H	494.45	3,291.34	187,187.341936	38,877.06	-109,433.220060
I	789.90	5,723.14	173,873.151720	43,044.57	-87,784.019292
J	826.10	10,992.00	172,241.840172	43,555.19	-85,131.460894
K	864.30	11,590.00	170,520.403443	44,094.02	-82,332.354836
L	869.30	15,973.00	170,295.084698	44,164.55	-81,965.979897
M	976.50	7,970.67	165,464.262738	45,676.67	-74,110.913120
N	1,355.80	9,524.10	148,371.862469	51,026.94	-46,317.989913
O	1,360.90	35,927.22	148,142.044389	51,098.87	-45,944.294515
P	1,463.21	11,238.73	143,531.737673	52,542.02	-38,447.695941
Q	2,332.10	92,059.97	104,382.272484	64,798.25	25,214.232669
R	3,017.73	74,649.00	75,911.693364	74,469.49	73,027.283557
S	3,253.00	42,915.23	92,744.870060	77,788.12	62,831.365052
Sums	19,633.77	542,585.74		883,094.70	

Table 29. Eighty Percent Upper and Lower Prediction Bounds for Adaptive-CER-Based Estimates at Cost-Driver Value 3,000

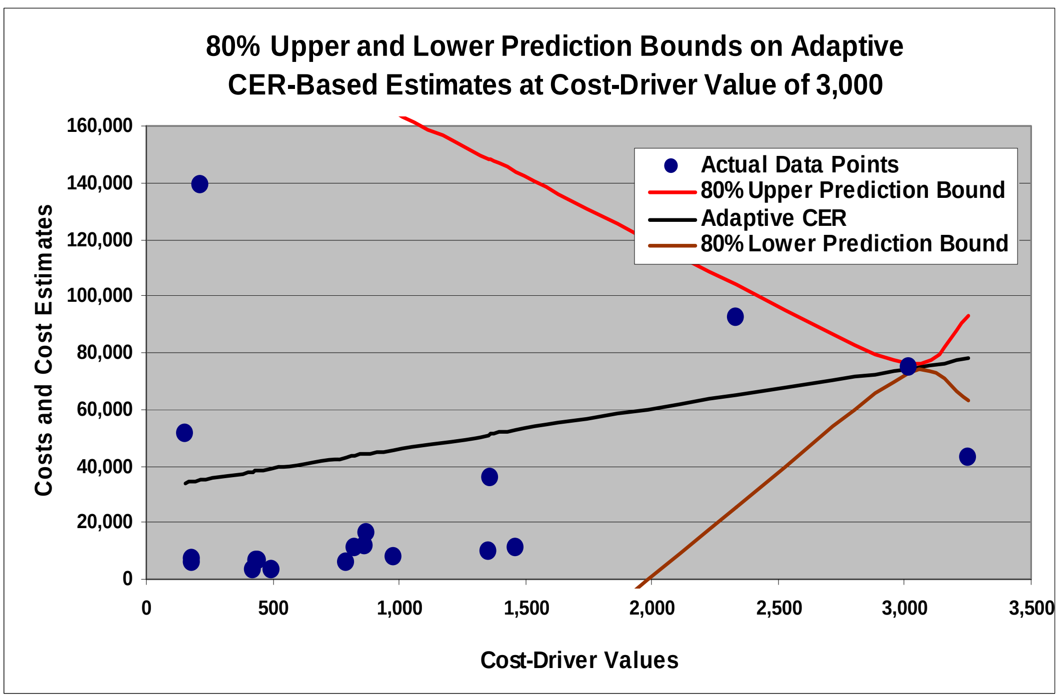

Figure 11. Eighty Percent Prediction Bounds for Adaptive-CER-Based Estimates at Cost-Driver Value 3,000 with Actual Data Points and Adaptive CER

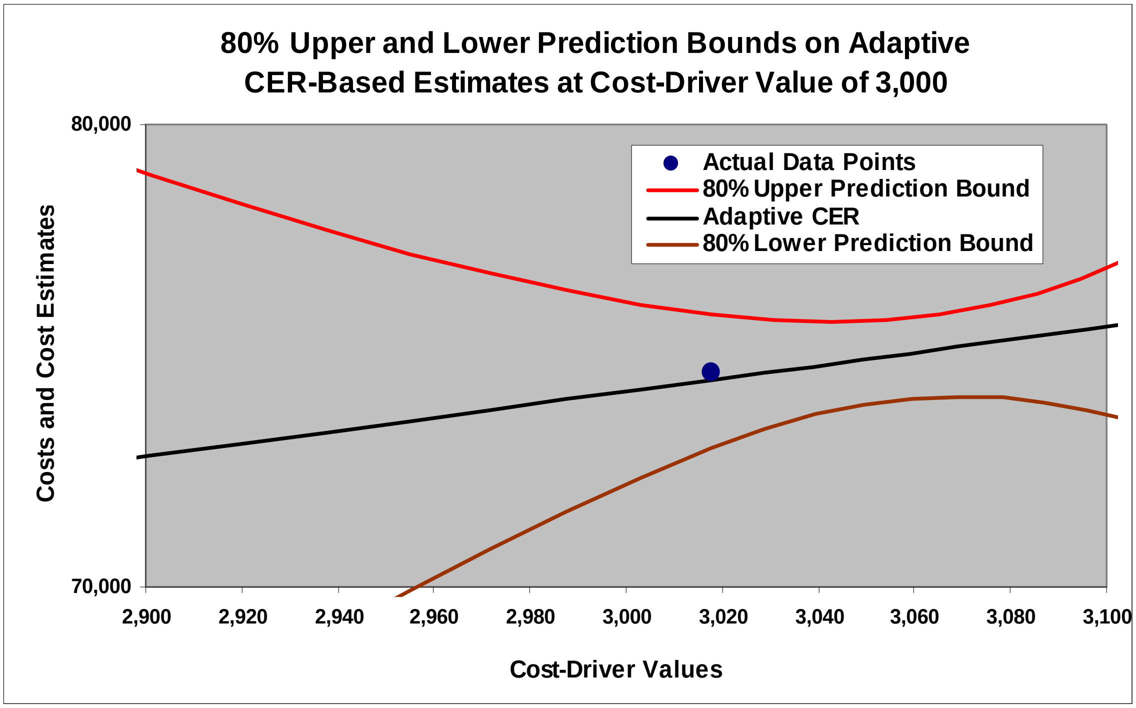

Figure 12. Gap between Upper and Lower Prediction Bounds in the Vicinity of the Cost-Driver Value 3,000 Prediction Bounds for the Universal Adaptive CER

The universal adaptive CER described in Table 25 and Figure 5 is formed by combining the various individual adaptive CERs, over the range of cost drivers into one CER that applies over the entire range. In the example we have been working with, adaptive CERs corresponding to 50-unit cost-driver increments are merged to form one continuous CER across the entire cost-driver range. The resulting universal adaptive CER is illustrated in Figure 5. Insofar as predictibounds are concerned, we want to make use of the fact that prediction bounds on each individual adaptive CER are very narrow in the vicinity of the cost-driver value on which the adaptive CER is based, but they widen considerably as the cost-driver value moves away from that point. This effect can be seen very clearly in Figures 7, 9, and 11. The universal adaptive CER takes advantage of this situation by providing estimates that have the narrowest possible prediction bounds for all cost-driver values.

Table 30 contains the numerical data on 80% upper and lower prediction bounds on estimate made using the universal adaptive CER. The prediction bounds themselves, along with the data points and the CER, appear in Figure 10. Note, that the prediction bounds are much narrower in the adaptive context than in the standard least-squares-fit context.

Driver	Cost	80% Upper Bound	EST Cost	80% Lower Bound
50.00		62,922.60536	42,739.31	22,556.01954
100.00		58,907.24807	40,817.29	22,727.33210
150.00		56,054.74733	49,546.82	43,038.89123
156.12	51,367.22	59,905.78998	50,880.53	41,855.27867
179.40	5,885.00	74,603.67051	55,953.88	37,304.09301
180.30	7,060.00	75,171.88754	56,150.02	37,128.14511
200.00		87,612.53844	60,443.18	33,273.82964
217.50	139,483.12	97,311.65891	69,749.17	42,186.69003
250.00		115,347.90219	87,031.73	58,715.55405
300.00		71,377.71021	46,425.71	21,473.71561
350.00		38,704.87919	22,733.56	6,762.24433
400.00		12,204.28249	7,006.95	1,809.62688
419.14	3,386.00	10,701.37240	6,760.42	2,819.47622
437.09	6,738.00	9,303.25537	6,529.22	3,755.18780
440.93	6,812.00	9,004.15958	6,479.76	3,955.36231
450.00		8,300.59270	6,362.94	4,425.27919
494.45	3,291.34	4,923.86478	3,589.46	2,255.05196
500.00		4,503.97498	3,243.16	1,982.35231
550.00		14,529.66385	6,829.12	-871.42873
600.00		19,484.26578	9,959.40	434.52824
650.00		20,409.64947	11,310.17	2,210.70010
700.00		18,067.87906	10,929.01	3,790.13759
750.00		12,749.77204	8,652.67	4,555.56975
789.90	5,723.14	9,150.40455	7,175.24	5,200.06839
800.00		8,241.00254	6,801.25	5,361.49607
826.10	10,992.00	11,221.66628	9,756.59	8,291.51518
850.00		13,951.60979	12,462.82	10,974.03604
864.30	11,590.00	14,422.75320	12,666.50	10,910.25030
869.30	15,973.00	14,587.63569	12,737.72	10,887.80057
900.00		15,604.93947	13,174.99	10,745.03389
950.00		11,653.20568	9,208.15	6,763.08930
976.50	7,970.67	11,143.81693	8,832.68	6,521.53760
1,000.00		10,696.45067	8,499.71	6,302.97553
1,050.00		16,599.85083	11,462.16	6,324.47866
1,100.00		20,939.42063	14,296.49	7,653.56932
1,150.00		23,860.71099	16,537.15	9,213.58728
1,200.00		25,595.54200	18,230.99	10,866.44026
1,250.00		26,430.13239	19,495.31	12,560.49388
1,300.00		26,643.54266	20,310.23	13,976.92318
1,350.00		19,965.36192	14,974.31	9,983.26614
1,355.80	9,524.10	20,888.41315	15,774.27	10,660.11771
1,360.90	35,927.22	21,705.81036	16,477.67	11,249.53159
1,400.00		27,979.59574	21,870.45	15,761.29785
1,450.00		13,664.60151	11,840.86	10,017.11120
1,463.21	11,238.73	14,382.93075	12,101.01	9,819.08646

Driver	Cost	80% Upper Bound	EST Cost	80% Lower Bound
1,500.00		16,394.47396	12,825.54	9,256.59722
1,550.00		22,698.80390	16,621.72	10,544.64424
1,600.00		28,144.34523	20,492.26	12,840.17489
1,650.00		33,397.70393	24,526.56	15,655.41028
1,700.00		38,715.42463	28,831.03	18,946.63797
1,750.00		44,154.10037	33,415.50	22,676.89870
1,800.00		49,694.94147	38,247.16	26,799.37082
1,850.00		55,295.14895	43,285.50	31,275.84370
1,900.00		60,905.74386	48,497.85	36,089.94751
1,950.00		66,472.33508	53,862.57	41,252.81236
2,000.00		71,925.95418	59,364.10	46,802.23932
2,050.00		77,167.60488	64,981.01	52,794.40890
2,100.00		82,049.60587	70,666.52	59,283.42427
2,150.00		86,358.35570	76,319.27	66,280.18315
2,200.00		89,805.61668	81,744.09	73,682.56956
2,250.00		92,036.69169	86,609.89	81,183.08854
2,300.00		92,664.98297	90,430.47	88,195.96100
2,332.10	92,059.97	93,449.79687	91,836.14	90,222.47807
2,350.00		93,889.12293	92,619.98	91,350.84125
2,400.00		97,402.84769	92,676.25	87,949.65291
2,450.00		98,222.52317	90,463.37	82,704.22441
2,500.00		96,484.73846	86,410.39	76,336.03984
2,550.00		92,951.10708	81,412.53	69,873.94518
2,600.00		88,646.33546	76,466.46	64,286.59020
2,650.00		84,454.68294	72,322.92	60,191.16611
2,700.00		80,929.09901	69,366.76	57,804.41474
2,750.00		77,054.82704	66,431.86	55,808.89003
2,800.00		76,663.27434	67,242.40	57,821.52197
2,850.00		75,905.67799	67,904.22	59,902.76737
2,900.00		75,867.66447	69,545.45	63,223.22554
2,950.00		76,119.31586	71,913.26	67,707.21079
3,000.00		75,518.29497	74,219.40	72,920.49668
3,017.73	74,649.00	75,966.58830	74,164.83	72,363.08019
3,050.00		76,786.35756	74,065.53	71,344.69813
3,100.00		74,002.10190	67,141.02	60,279.92945
3,150.00		62,069.86209	54,415.99	46,762.11593
3,200.00		49,725.94543	45,424.15	41,122.36282
3,250.00		43,144.36743	42,927.10	42,709.82647
3,253.00	42,915.23	43,354.47526	42,978.65	42,602.82964
3,300.00		46,653.41208	43,786.36	40,919.29842
3,350.00		50,744.79550	45,762.39	40,779.98882
3,400.00		54,430.68793	47,971.96	41,513.24151
3,450.00		57,664.17277	50,126.95	42,589.72220
3,500.00		60,512.13570	52,149.51	43,786.89143

Table 30. Universal Adaptive-CER-Based Estimates and 80% Prediction Bounds at 50-Unit Increments Along the Cost-Driver Axis

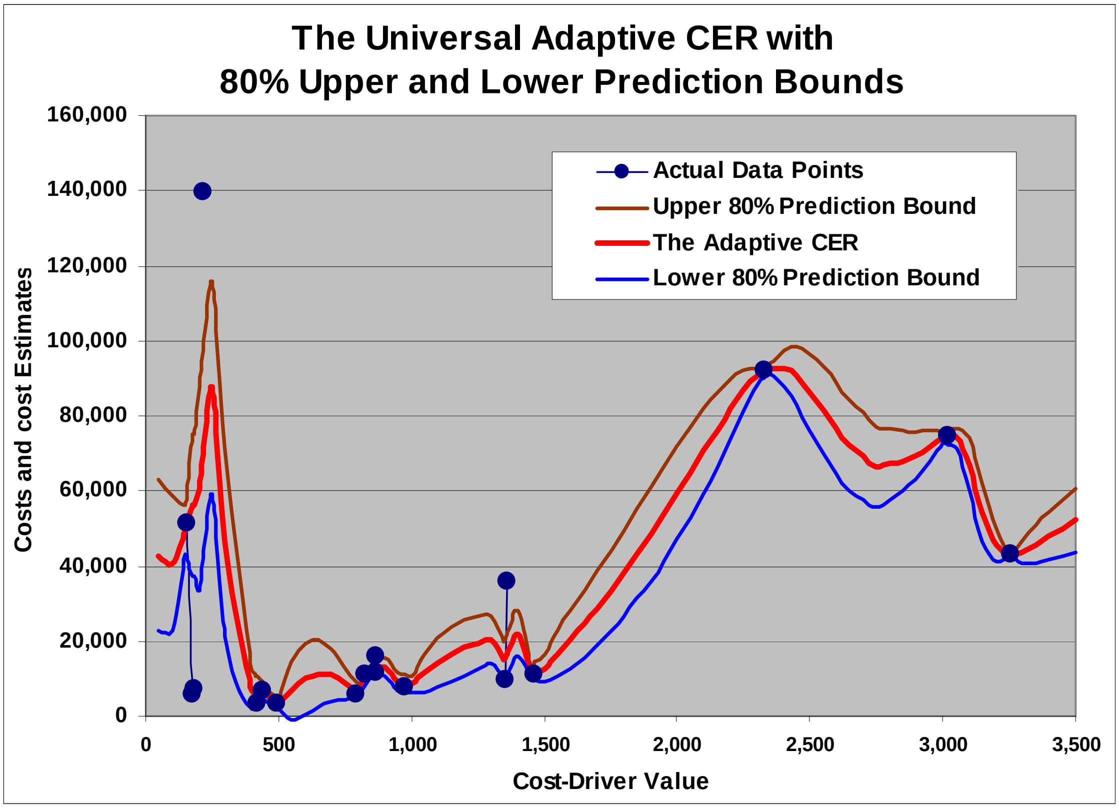

Figure 13. Universal Adaptive-CER-Based Estimates and 80% Prediction Bounds Graphed at 50-Unit Increments along the Cost-Driver Axis

As is characteristic of adaptive CERs, we see that the prediction bounds are much narrower in Figure 10 than they are in the OLS regression situation illustrated in Figure 6. Again, this narrowing is due to the fact that estimating using an adaptive CER near a cost-driver value is carried out using only data points near that cost-driver value. However, when there is significant variation in data points near a cost-driver value, the prediction bounds widen in that region. For an example, see what happens in the cost-driver region of 200-300 in Figure 13 above. The prediction bounds for OLS CERs, on the other hand, must be wide enough to provide the desired amount of confidence, e.g., 80%, throughout the entire cost-driver range.

Acknowledgements

The authors are grateful to the members of MCR's quality-review team: Dr. Neal D. Hulkower, Dr. Jerry D. Hofmann, Raymond P. Covert, Nathan J. Menton, Timothy P. Anderson, and Jan B. Sterbutzel. The combined efforts of the team resulted in a significant improvement in the clarity of explanation of specific details of the conclusions drawn, as well as the elimination of various typos and other errors.

Appendix

Algebraic Analysis of the Total Variation

$$
\begin{aligned}
TV &= \sum_{k=1}^{n} w_k (y_k - \bar{y}_w)^2 = \sum_{k=1}^{n} w_k \left[(y_k - a - bx_k) + (a + bx_k - \bar{y}_w)\right]^2 \\
&= \sum_{k=1}^{n} w_k \left[(y_k - a - bx_k)^2 + 2(y_k - a - bx_k)(a + bx_k - \bar{y}_w) + (a + bx_k - \bar{y}_w)^2\right] \\
&= \sum_{k=1}^{n} w_k (y_k - a - bx_k)^2 + \sum_{k=1}^{n} w_k (a + bx_k - \bar{y}_w)^2 + 2\sum_{k=1}^{n} w_k (y_k - a - bx_k)(a + bx_k - \bar{y}_w) \\
&= SS + VB + 2\sum_{k=1}^{n} w_k (y_k - a - bx_k)(a + bx_k - \bar{y}_w)
\end{aligned}
$$

We now show that the third summand in the above equation is always zero, no matter what the data, so that ***TV = SS + VB*** for every set of data points. The expression for ***a*** that results from solving for the WLS regression equation implies that

$$
a = \frac{\left(\sum_{k=1}^{n} w_k y_k\right)}{\left(\sum_{k=1}^{n} w_k\right)} - b\,\frac{\left(\sum_{k=1}^{n} w_k x_k\right)}{\left(\sum_{k=1}^{n} w_k\right)} = \bar{y}_w - b\bar{x}_w\,,
$$

where $\bar{y}_w$ and $\bar{x}_w$ are the weighted means of the ***y*** and ***x*** values in the data set, respectively. Therefore $a + bx_k - \bar{y}_w = a + bx_k - (a + b\bar{x}_w) = b(x_k - \bar{x}_w)$, from which it follows that

$$2\sum_{k=1}^{n} w_k(y_k - a - bx_k)(a + bx_k - \bar{y}_w) = 2\sum_{k=1}^{n} w_k(y_k - a - bx_k)b(x_k - \bar{x}_w)$$

$$= 2b\sum_{k=1}^{n} w_k(x_k y_k - ax_k - bx_k^2 - \bar{x}_w y_k + a\bar{x}_w + b\bar{x}_w x_k)$$

$$= 2b\left[\sum_{k=1}^{n} w_k x_k y_k - a\sum_{k=1}^{n} w_k x_k - b\sum_{k=1}^{n} w_k x_k^2 - \bar{x}_w\sum_{k=1}^{n} w_k y_k + a\bar{x}_w\sum_{k=1}^{n} w_k + b\bar{x}_w\sum_{k=1}^{n} w_k x_k\right]$$

In view of the fact that $\sum_{k=1}^{n} w_k x_k = \bar{x}_w\sum_{k=1}^{n} w_k$, the two terms above that contain "***a***" can be canceled out. What remains is, except for the "***2b***" factor:

$$\sum_{k=1}^{n} w_k x_k y_k - b\sum_{k=1}^{n} w_k x_k^2 - \bar{x}_w\sum_{k=1}^{n} w_k y_k - b\bar{x}_w\sum_{k=1}^{n} w_k x_k$$

$$= \sum_{k=1}^{n} w_k x_k y_k - \frac{\left(\sum_{k=1}^{n} w_k x_k\right)\left(\sum_{k=1}^{n} w_k y_k\right)}{\sum_{k=1}^{n} w_k} - b\sum_{k=1}^{n} w_k x_k^2 - b\frac{\left(\sum_{k=1}^{n} w_k x_k\right)^2}{\sum_{k=1}^{n} w_k}$$

$$= \frac{\sum_{k=1}^{n} w_k \sum_{k=1}^{n} w_k x_k y_k - \left(\sum_{k=1}^{n} w_k x_k\right)\left(\sum_{k=1}^{n} w_k y_k\right)}{\sum_{k=1}^{n} w_k} - b\left[\frac{\sum_{k=1}^{n} w_k \sum_{k=1}^{n} w_k x_k^2 - \left(\sum_{k=1}^{n} w_k x_k\right)^2}{\sum_{k=1}^{n} w_k}\right]$$

$$= \frac{b\left[\left(\sum_{k=1}^{n} w_k\right)\left(\sum_{k=1}^{n} w_k x_k^2\right) - \left(\sum_{k=1}^{n} w_k x_k\right)^2\right]}{\sum_{k=1}^{n} w_k} - b\left[\frac{\sum_{k=1}^{n} w_k \sum_{k=1}^{n} w_k x_k^2 - \left(\sum_{k=1}^{n} w_k x_k\right)^2}{\sum_{k=1}^{n} w_k}\right] = 0$$

Because $$b = \frac{\left(\sum_{k=1}^{n} w_k\right)\left(\sum_{k=1}^{n} w_k x_k y_k\right) - \left(\sum_{k=1}^{n} w_k x_k\right)\left(\sum_{k=1}^{n} w_k y_k\right)}{\left(\sum_{k=1}^{n} w_k\right)\left(\sum_{k=1}^{n} w_k x_k^2\right) - \left(\sum_{k=1}^{n} w_k x_k\right)^2}$$

References

1. Book, S.A., “Deriving Cost-Estimating Relationships Using Weighted Least-Squares Regression,” IAA/ISPA/AIAA Space System Cost Methodologies and Applications Symposium, San Diego CA, 10-11 May 1990, 19 pages.

2. Book, S.A.and Broder, M.A., “Adaptive Cost-Estimating Relationships,” Space Systems Cost Analysis Group (SSCAG) Meeting in Noordwijk, Holland, 15-16 May 2008, and ISPA/SCEA Joint Annual Conference & Training Workshop, Industry Hills CA, 24-27 June 2008, 80 Charts.

3. Foussier, P., “Space System Study,” Space Systems Cost Analysis Group (SSCAG), Montreal QB, 16 May 2007, 41 Charts.

Author Biographies

Dr. Stephen A. Book is Chief Technical Officer of MCR, LLC. In that capacity, he is responsible for ensuring technical excellence of MCR’s products, services, and processes by encouraging process improvement, maintaining quality control, and training employees and customers in cost and schedule analysis and associated program-control disciplines. During his career, he has provided technical support in the cost, schedule, and earned-value areas to several Air Force, NASA, and Intelligence organizations and continues to do so. Dr. Book joined MCR in January 2001 after 21 years with The Aerospace Corporation, where he held the title “Distinguished Engineer” during 1996-2000 and served as Director, Resource and Requirements Analysis Department, during 1989-1995. Dr. Book was the last editor of ISPA’s *Journal of Parametrics* prior to its merger with SCEA’s *Journal of Cost Analysis and Management*, and is now co-editor of the combined journal. He was the 2005 recipient of ISPA’s Freiman Award for Lifetime Achievement. He earned his Ph.D. in mathematics, with concentration in probability and statistics, at the University of Oregon.

Melvin A. Broder is a Senior Project Leader at The Aerospace Corporation. In that capacity he has developed cost models for the Concept Design Center, building and expanding tools for the cost seat and devising new processes. Prior to joining Aerospace he worked in cost estimating at Boeing's Satellite Systems where he was responsible for front end of the business cost tools and models for Boeing’s commercial product line

including support of the BSS Design Center's Integrated Engineering Laboratory. Mr. Broder has also been a Project Manager for cost tools and processes in the System Engineering Laboratory at Raytheon Systems Company. His responsibilities include the creation and maintenance of tools to support the Sensors and Electronic Systems in design to cost (DTC) and cost-as-an-independent-variable (CAIV) activities. Prior to working the aerospace industry he was an Instructor of Economics at La Verne College, teaching both upper and lower division course work in Micro and Macro Economics, Comparative Economics, Money and Banking, and Econometrics. He earned an M.S. in Economics from the University of Southern California.

Daniel I. Feldman is a Junior Cost Analyst at MCR, LLC. Since joining MCR in early September 2005, he has worked on developing new techniques in utilizing CER-based estimates, along with doing rocket modeling and trade-study analysis. Mr. Feldman earned his B.S. in mathematics in June 2005, with concentration in statistics, at the University of California, Irvine. He is currently working on his M.S. in applied statistics at California State University, Long Beach.

APPENDIX B. LI

A. LINEAR REGRESSION WITH TWO VARIABLES

In order to find more specific cost drivers, two variable linear regressions were examined with average unit costs and 36 combinations of 2 variables from 9 cost driver factors. The results appear in Table 22.

Table 31. The Result of Linear Regression with Two Variables

Y	Independent variable		Linear Regression with two variables						
	X_1	X_2	P-value				Significance F	R Square	Estimate
Average Unit Cost	Max Taking-Off	Max disc loading	X_1	0.000449	X_2	0.013485	0.000381	0.957119	9.324
			Equation	$y= -4.2752+ 0.000621X_1+0.218676X_2$					
		SHP	X_1	0.200773	X_2	0.066932	0.001690	0.922164	13.855
			Equation	$y= 1.625232 - 0.001242 X_1+0.006288X_2$					
		Main Rotor	X_1	0.235625	X_2	0.279964	0.005603	0.874287	12.800
			Equation	$y= -3.616629+ 0.000393 X_1+0.817794X_2$					
		Height	X_1	0.206918	X_2	0.658838	0.009572276	0.844257881	11.385
			Equation	$y= -1.383963+ 0.000547 X_1+1.770924X_2$					
		Max speed	X_1	0.002886	X_2	0.389015	0.007080297	0.861955031	11.360
			Equation	$y= 1.084886+ 0.000704 X_1+0.013368X_2$					
		Cruising speed	X_1	0.004285	X_2	0.847122	0.010443257	0.83873713	11.390
			Equation	$y= 2.995083+ 0.000707 X_1+0.009031X_2$					
		Max Range	X_1	0.000369	X_2	0.037046	0.000985421	0.937273832	14.114
			Equation	$y= 13.670328+ 0.000751X_1 - 0.013928X_2$					
		Empty Weight	X_1	0.506888	X_2	0.522173	0.008503438	0.851461915	11.808
			Equation	$y= 4.532604+ 0.000369X_1+0.000808X_2$					

Y	Independent variable		Linear regression with 2 variables						
	X_1	X_2	P-value				Significance F	R Square	Estimate
Average Unit Cost	Max disc loading	SHP	X_1	0.011182	X_2	0.000145	0.00012427	0.9726	10.381
			Equation	$y= -4.133102+ 0.187267X1+0.002054X_2$					
		Main Rotor	X_1	0.303087	X_2	0.008495	0.00677693	0.864352	12.369
			Equation	$y= -15.230423+ 0.130914X1+1.443643X_2$					
		Height	X_1	0.100292	X_2	0.006479	0.005217833	0.877821	8.749
			Equation	$y= -24.908503+ 0.203015X1+5.884112X_2$					
		Max speed	X_1	0.113592	X_2	0.473062	0.222432088	0.451876	7.141
			Equation	$y= -2.87384+ 0.50944X1-0.02932X_2$					
		Cruising speed	X_1	0.153744	X_2	0.834093	0.288415893	0.391854	8.321
			Equation	$y= -1.4094762+ 0.396026X1-0.021075X_2$					
		Max Range	X_1	0.063765	X_2	0.246551	0.141046513	0.54318	0.912
			Equation	$y= -32.959891+ 0.61688X1+0.024811X_2$					
		Empty Weight	X_1	0.173005	X_2	0.004803	0.003903665	0.89121	10.315
			Equation	$y= -2.40047+ 0.157133X1+0.001408X_2$					

Y	Independent variable		Linear regression with 2 variables						
	X_1	X_2	P-value				Significance F	R Square	Estimate
Average Unit Cost	SHP	Main Rotor	X_1	0.128196	X_2	0.547106	0.003407	0.896976	12.893
			Equation	$y= -0.80331+ 0.001764X1+0.453228 X_2$					
		Height	X_1	0.077477	X_2	0.966216	0.004157214	0.888437	12.413
			Equation	$y= 3.228985+ 0.002302X1+0.144638 X_2$					
		Max speed	X_1	0.001222	X_2	0.452784	0.00304928	0.901445	12.241
			Equation	$y= 0.746511 + 0.002312X1 + 0.009788 X_2$					
		Cruising speed	X_1	0.001671	X_2	0.976621	0.004159361	0.888414	12.479
			Equation	$y= 4.032276 + 0.002348X1 - 0.001149 X_2$					
		Max Range	X_1	7.36E-05	X_2	0.018315	0.000198852	0.966932	14.678
			Equation	$y= 11.204742+ 0.002426X1 - 0.012283 X_2$					
		Empty Weight	X_1	0.18712	X_2	0.998308	0.004161324	0.888393	12.446
			Equation	$y= 3.746793+ 0.002342X1+ 0.000002 X_2$					
	Main Rotor	Height	X_1	0.276174	X_2	0.866748	0.011970256	0.82969	13.608
			Equation	$y= -13.150464+ 1.442118X1 + 0.899315 X_2$					
		Max speed	X_1	0.00492	X_2	0.845913	0.011906889	0.830051	13.933
			Equation	$y= -12.732+ 1.627871X1 + 0.00328 X_2$					
		Cruising speed	X_1	0.004611	X_2	0.702313	0.011215829	0.834067	18.835
			Equation	$y= -7.698531 + 1.686802X1 - 0.018877 X_2$					
		Max Range	X_1	0.003597	X_2	0.48285	0.009265585	0.846273	20.259
			Equation	$y= -8.101975+ 1.632198X1 - 0.005789 X_2$					
		Empty Weight	X_1	0.336727	X_2	0.287568	0.006519575	0.866437	12.993
			Equation	$y= -4.122732 + 0.807924X1 + 0.000887 X_2$					

Y	Independent variable		Linear regression with 2 variables						
	X_1	X_2	P-value				Significance F	R Square	Estimate
Average Unit Cost	Height	Max speed	X_1	0.004207	X_2	0.224932	0.010225329	0.840092	10.306
			Equation	$y= -26.779382+ 6.922776X1+ 0.021072 X_2$					
		Cruising speed	X_1	0.00916	X_2	0.752934	0.021780892	0.783613	10.310
			Equation	$y= -23.758330+ 6.775005X1+ 0.017040 X_2$					
		Max Range	X_1	0.007396	X_2	0.537135	0.018644783	0.79666	11.805
			Equation	$y= -15.950959 + 6.825813X1 - 0.005820 X_2$					
		Empty Weight	X_1	0.367728	X_2	0.139676	0.006932263	0.863117	11.513
			Equation	$y= -5.749472+ 2.683295X1+ 0.001081 X_2$					
	Max speed	Cruising speed	X_1	0.863593	X_2	0.877435	0.868912891	0.054655	11.027
			Equation	$y= 1.221152+ 0.011062X1+ 0.0283 X_2$					
		Max Range	X_1	0.666454	X_2	0.766536	0.838574308	0.067998	12.738
			Equation	$y= 10.34775 + 0.017044X1 - 0.005975 X_2$					
		Empty Weight	X_1	0.699661	X_2	0.004106	0.00998762	0.841589	12.138
			Equation	$y= 5.907661 - 0.006533X1+ 0.001661 X_2$					
	Cruising speed	Max Range	X_1	0.651439	X_2	0.73766	0.830143201	0.071758	11.865
			Equation	$y= 3.284156+ 0.050353X1 - 0.006667 X_2$					
		Empty Weight	X_1	0.459237	X_2	0.003266	0.008015278	0.854933	12.959
			Equation	$y= 12.741776 - 0.035787X1 + 0.001716 X_2$					
	Max Range	Empty Weight	X_1	0.050714	X_2	0.000502	0.001337536	0.92912	14.423
			Equation	$y= 12.10339 - 0.0134X1 + 0.001696 X_2$					

B. POWER REGRESSION WITH TWO VARIABLES

In order to find more specific cost drivers and to fit the non-linear to linear, nine variables power regressions were examined with average unit costs and 36 combinations of two variables from nine cost driver factors which the results displayed in Table 23. This is the whole result of power regression with two variables.

Table 32. The Result of Power Regression with Two Variables

Y	Independent variable		Power regression with two variables						
	X_1	X_2	P-value				Significance F	R Square	Estimate
Average Unit Cost	Max Taking-Off	Max disc loading	X_1	0.000843	X_2	0.050510	0.000279	0.962152	9.896
			Equation	$y=0.005459*X_1^{0.540361}* X_2^{0.717311}$					
		SHP	X_1	0.475732	X_2	0.154943	0.000745	0.943904	13.435
			Equation	$y=0.022817* X_1^{(-0.567253)}* X_2^{1.403961}$					
		Main Rotor	X_1	0.224763	X_2	0.937747	0.002260	0.912575	12.025
			Equation	$y=0.029827* X_1^{0.622830}* X_2^{0.120689}$					
		Height	X_1	0.026838	X_2	0.470543	0.001701368	0.92196	12.601
			Equation	$y=0.022342*X_1^{0.866399}* X_2^{(-1.036432)}$					
		Max speed	X_1	0.000958	X_2	0.566705	0.001891828	0.918576	11.778
			Equation	$y=0.010658*X_1^{0.641529}* X_2^{0.205566}$					
		Cruising speed	X_1	0.001025	X_2	0.821671	0.002204786	0.913434	12.190
			Equation	$y=0.080076*X_1^{0.667586}* X_2^{(-0.192745)}$					
		Max Range	X_1	0.000168	X_2	0.074604	0.000396195	0.956432	13.787
			Equation	$y=0.599107*X_1^{0.648372}*X_2^{(-0.452229)}$					
		Empty Weight	X_1	0.219016	X_2	0.770187	0.002163145	0.914092	11.994
			Equation	$y=0.030753*X_1^{0.543078}* X_2^{0.120606}$					

Y	Independent variable		Power regression with two variables							
	X_1	X_2	P-value				Significance F	R Square	Estimate
Average Unit Cost	Max disc loading	SHP	X_1	0.039815	X_2	0.000292	9.74929E-05	0.975135	10.667
			Equation	$y=0.005687*X_1^{0.637706}*X_2^{0.637240}$					
		Main Rotor	X_1	0.110885	X_2	0.003923	0.001258872	0.930818	10.988
			Equation	$y=0.006873*X_1^{0.738957}*X_2^{1.708173}$					
		Height	X_1	0.021639	X_2	0.004343	0.001389817	0.928025	7.872
			Equation	$y=0.004887*X_1^{1.138692}*X_2^{2.196166}$					
		Max speed	X_1	0.047981	X_2	0.422891	0.081408153	0.633337	6.301
			Equation	$y=0.077842*X_1^{2.535156}*X_2^{(-0.833310)}$					
		Cruising speed	X_1	0.062627	X_2	0.829612	0.113084256	0.581822	7.134
			Equation	$y=0.049567*X_1^{2.033087}*X_2^{(-0.434250)}$					
		Max Range	X_1	0.024273	X_2	0.221182	0.050888004	0.696159	4.008
			Equation	$y=0.00000033*X_1^{2.781387}*X_2^{1.028496}$					
		Empty Weight	X_1	0.329439	X_2	0.009431	0.002942188	0.902844	10.601
			Equation	$y=0.014046*X_1^{0.533187}*X_2^{0.553286}$					

Y	Independent variable		Power regression with two variables						
	X_1	X_2	P-value				Significance F	R Square	Estimate
Average Unit Cost	SHP	Main Rotor	X_1	0.071038	X_2	0.604476	0.000850888	0.940851	12.408
			Equation	$y=0.023037* X_1^{0.997235}* X_2^{(-0.691288)}$					
		Height	X_1	0.010964	X_2	0.457913	0.000728088	0.944426	13.481
			Equation	$y=0.019516* X_1^{0.941012}* X_2^{(-0.801264)}$					
		Max speed	X_1	0.00043	X_2	0.614342	0.000857673	0.940663	12.590
			Equation	$y=0.011423*X_1^{0.745181}* X_2^{0.154558}$					
		Cruising speed	X_1	0.000405	X_2	0.650467	0.000880976	0.940023	13.226
			Equation	$y=0.130930* X_1^{0.778218}* X_2^{(-0.327457)}$					
		Max Range	X_1	3.8E-05	X_2	0.036793	9.06362E-05	0.975850	14.561
			Equation	$y=0.417159*X_1^{0.747558}* X_2^{(-0.424173)}$					
		Empty Weight	X_1	0.085834	X_2	0.937586	0.000983362	0.937326	12.747
			Equation	$y=0.024194* X_1^{0.790503}* X_2^{(-0.027061)}$					
	Main Rotor	Height	X_1	0.085344	X_2	0.789498	0.004891125	0.880941	14.298
			Equation	$y=0.036333* X_1^{2.416261}* X_2^{(-0.462642)}$					
		Max speed	X_1	0.002134	X_2	0.544793	0.00415537	0.888457	13.978
			Equation	$y=0.01164* X_1^{2.046373}* X_2^{0.253706}$					
		Cruising speed	X_1	0.00237	X_2	0.879752	0.005023544	0.879662	25.417
			Equation	$y=0.088074* X_1^{2.135901}* X_2^{(-0.15233)}$					
		Max Range	X_1	0.000911	X_2	0.207021	0.002117637	0.914819	15.543
			Equation	$y= 0.621106* X_1^{2.072174}* X_2^{(-0.408678)}$					
		Empty Weight	X_1	0.382741	X_2	0.369797	0.003265114	0.898712	12.974
			Equation	$y=0.038486* X_1^{1.064206}* X_2^{0.339291}$					

Y	Independent variable		Power regression with two variables						
	X_1	X_2	P-value				Significance F	R Square	Estimate
Average Unit Cost	Height	Max speed	X_1	0.002428	X_2	0.081871	0.004714772	0.882677	9.729
			Equation	$y=0.001226*X_1^{2.836414}* X_2^{0.832812}$					
		Cruising speed	X_1	0.00976	X_2	0.523401	0.019845278	0.791521	9.334
			Equation	$y=0.001690*X_1^{2.780785}* X_2^{0.821112}$					
		Max Range	X_1	0.006883	X_2	0.347899	0.015285743	0.812192	11.711
			Equation	$y= 2.449265* X_1^{2.821643}* X_2^{(-0.433390)}$					
		Empty Weight	X_1	0.38723	X_2	0.054841	0.003290846	0.898393	11.631
			Equation	$y=0.047325* X_1^{0.879271})* X_2^{0.493048}$					
	Max speed	Cruising speed	X_1	0.677339	X_2	0.899549	0.688109661	0.138881	10.171
			Equation	$y=0.008954* X_1^{0.744857}* X_2^{0.513357}$					
		Max Range	X_1	0.470009	X_2	0.633266	0.612215251	0.178208	12.394
			Equation	$y= 1.950697* X_1^{0.810226}* X_2^{(-0.452909)}$					
		Empty Weight	X_1	0.407553	X_2	0.001741	0.003404568	0.897003	12.594
			Equation	$y=0.261622* X_1^{(-0.377468)}* X_2^{0.708633}$					
	Cruising speed	Max Range	X_1	0.502149	X_2	0.575559	0.636845445	0.16514	11.050
			Equation	$y= 0.026684*X_1^{1.702750}* X_2^{(-0.529293)}$					
		Empty Weight	X_1	0.245471	X_2	0.0011	0.002364279	0.910982	14.025
			Equation	$y=16.5484*X_1^{(-1.166667)}* X_2^{0.726801}$					
	Max Range	Empty Weight	X_1	0.137023	X_2	0.000631	0.001473531	0.926321	14.198
			Equation	$y=1.004258*X_1^{(-0.463106)}* X_2^{0.644349}$					

C. WEIGHTED DATA

Historical data on helicopter development is difficult to obtain, either because of security or proprietary concerns. Instead, the author collected data from open sources. The main source of data was *Jane's All the World's Aircraft.* Other data sources are listed in the Reference section. After then, weight was assigned to each cost driver as mentioned in III.D.1.

Table 24 displays the data collected for this thesis. There are eight helicopters, each with nine descriptive variables.

Table 33. Weight Assignment and Weighted Data

Name	Type	Weight			Average Unit cost($M)	Weight(kg)			Power Plant (SHP)	Dimensions(M)		Speed (km/h)	Max Range
		Initial weight * penalty	normalized weight	sqrt(normalized weight)		Empty	Max Taking-Off	Max disc loading (kg/m2)		Main Rotor	Height	Max speed	km
UH-1Y	Medium Utility	9	0.1500	0.3873	4.40	2,079.79	3,249.43	19.33	1,197.53	5.67	1.72	141.75	265.69
AH-1Z	Attack	7.2	0.1200	0.3464	3.91	1,932.97	2,907.07	17.29	1,193.73	5.06	1.51	142.37	237.64
CH-47D	Cargo	3.2	0.0533	0.2309	4.66	2,344.27	5,237.72	10.85	1,732.05	4.30	1.32	68.82	171.13
AH-64	Attack	7.2	0.1200	0.3464	5.27	1,789.21	3,299.56	21.51	1,247.08	5.07	1.61	126.44	140.99
EC-145	Utility	7.2	0.1200	0.3464	2.21	624.92	1,241.88	13.06	533.47	3.81	1.37	92.84	235.56
AS - 532UB	Medium Utility	10	0.1667	0.4082	5.76	1,767.72	3,674.23	19.96	1,532.56	6.37	1.96	113.49	233.93
UH-60L	Utility	9	0.1500	0.3873	4.46	2,023.25	4,128.60	18.28	1,463.99	6.35	2.01	113.87	226.18
UH-72A LAKOTA	Utility	7.2	0.1200	0.3464	2.10	620.77	1,241.88	13.06	511.30	3.81	1.37	92.84	237.29

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Professor D. Nussbaum
   Naval Postgraduate School
   Monterey California

4. Professor Dr.Kang, Sungjin
   Department of Operations Research
   Korea National Defense University
   Seoul, Korea

5. Professor Dr. Lee, Jaeyoung
   Department of Operations Research
   Korea National Defense University
   Seoul, Korea